

GenCore version 5.1.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 3, 2003, 22:52:46 ; Search time 225.158 Seconds
(without alignments)
3231.380 Million cell updates/sec

Title: US-09-787-562-1
Perfect score: 25
Sequence: 1 cgcgcggcgagcagcgagcaaat 25

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

GenEmbl: *
1: gb.ba: *
2: gb.htg: *
3: gb.in: *
4: gb.om: *
5: gb.ov: *
6: gb.pat: *
7: gb.ph: *
8: gb.pl: *
9: gb.pr: *
10: gb.ro: *
11: gb.sts: *
12: gb.sy: *
13: gb.un: *
14: gb.vi: *
15: em.ba: *
16: em.fun: *
17: em.in: *
18: em.in: *
19: em.mu: *
20: em.mu: *
21: em.or: *
22: em.ov: *
23: em.pat: *
24: em.ph: *
25: em.pl: *
26: em.ro: *
27: em.sts: *
28: em.un: *
29: em.vi: *
30: em.htg_hum: *
31: em.htg_inv: *
32: em.htg_other: *
33: em.htg_mus: *
34: em.htg_pln: *
35: em.htg_rod: *
36: em.htg_mam: *
37: em.htg_vrt: *
38: em_sy: *
39: em_htgo_hum: *
40: em_htgo_mus: *
41: em_htgo_other: *

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
1	25	100.0	25	6	AX023659	Sequence
2	19.2	76.8	273785	1	SME591793	AL591793 Sinorhizo
3	19.2	76.8	338579	1	AP003004	AP003004 Mesorhizo
4	19	76.0	19	6	AX023660	Sequence
5	18.6	74.4	695	8	HOS271030	AJ271030 Hansenias
6	18.6	74.4	704	8	HVS271033	AJ271033 Hansenias
7	18.6	74.4	3670	1	VFU41417	U41417 Vibrio furn
8	18.6	74.4	3670	6	AR022345	Sequence
9	18.6	74.4	3670	6	AR086949	Sequence
10	18.6	74.4	217812	2	AC130788	AC130788 Bos tauru
11	18.4	73.6	117962	2	AP003914	AP003914 Oryza sat
12	18.4	73.6	154255	2	AP004632	AP004632 Oryza sat
13	18	72.0	24	6	A46287	Sequence 2
14	18	72.0	24	6	AX023673	Sequence
15	18	72.0	24	6	AX048713	Sequence
16	18	72.0	366	5	AF353996	Cyprinus
17	18	72.0	513	6	AX150246	Sequence
18	18	72.0	1110	10	MUSFGK1	M18735 Mouse phosph
19	18	72.0	1943	10	MMPGK5R	X15339 M.musculus
20	18	72.0	3297	12	AF090453	Cloning v
21	18	72.0	3426	12	AF090454	Cloning v
22	18	72.0	4699	12	AF346623	Cloning v
23	18	72.0	4768	6	AX299821	Sequence
24	18	72.0	4768	6	AX352704	Sequence
25	18	72.0	4847	6	AX191674	Sequence
26	18	72.0	5365	6	AX114854	Sequence
27	18	72.0	5482	12	AF090455	Sequence
28	18	72.0	5505	2	AL831717	Cloning v
29	18	72.0	5568	12	AF335419	Cloning v
30	18	72.0	5568	12	AF335420	Cloning v
31	18	72.0	5608	12	AF092169	Cloning v
32	18	72.0	5608	12	AF092172	Cloning v
33	18	72.0	5608	12	AF092173	Cloning v
34	18	72.0	5608	12	AF092174	Cloning v
35	18	72.0	5608	12	AF092541	Cloning v
36	18	72.0	5608	12	AF092542	Cloning v
37	18	72.0	5608	12	AF092543	Cloning v
38	18	72.0	5608	12	AF092567	Cloning v
39	18	72.0	5830	12	AY028413	Cloning v
40	18	72.0	5842	12	AY028415	YTT vecto
41	18	72.0	6293	12	AF397196	YTT vecto
42	18	72.0	6355	6	AX299822	Retrofitt
43	18	72.0	6355	6	AX352705	Sequence
44	18	72.0	7090	6	AX150263	Sequence
45	18	72.0	8157	2	AL831743	Mus muscu

ALIGNMENTS

RESULT 1	AX023659	Sequence 1 from Patent WO0017371.	25 bp	DNA	linear	PAT 15-SEP-2000
LOCUS	AX023659					
DEFINITION	AX023659					
ACCESSION	AX023659					
VERSION	AX023659.1	GI:10184020				
KEYWORDS						
SOURCE	Mus sp.					
ORGANISM	Mus sp.					
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;					
AUTHORS	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.					
TITLE	1 (bases 1 to 25)					
JOURNAL	Binley, K.M. and Naylor, S.					
	Polynucleotide constructs and uses thereof					
	Patent: WO 0017371-A 1 30-MAR-2000;					

BINLEY KATIE MARY (GB) ; NAYLOR STUART (GB) ; OXFORD BIOMEDICA LTD (GB)

FEATURES source Location/Qualifiers

1..25
/organism="Mus sp."
/db_xref="taxon:10095"

BASE COUNT 6 a 6 c 9 g 4 t

ORIGIN

Query Match 100.0%; Score 25; DB 6; Length 25;

Best Local Similarity 100.0%; Pred. No. 2.2; Mismatches 0; Indels 0; Gaps 0;

Matches 25; Conservative 0;

Qy 1 CGCGTCGGTCGAGCAGCGTGCACAAAT 25

|||||

Db 1 CGCGTCGGTCGAGCAGCGTGCACAAAT 25

RESULT 2

SME591793

LOCUS SME591793 273785 bp DNA linear BCT 05-JUL-2002

DEFINITION Sinorhizobium meliloti 1021 complete chromosome; segment 12/12.

ACCESSION AL591793 AL591688

VERSION AL591793.1 GI:15076142

KEYWORDS

SOURCE

ORGANISM

Sinorhizobium meliloti.

Sinorhizobium meliloti

Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;

Rhizobiaceae; Sinorhizobium.

1 (bases 1 to 273785)

Capela,D., Barloy-Hubier,F., Gouzy,J., Bothe,G., Ampe,F., Batut,J.,

Boistard,P., Becker,A., Boutry,M., Cadieu,E., Dreano,S., Gloux,S.,

Godrie,T., Goffeau,A., Kahn,D., Kiss,E., Lelaure,V., Masuy,D.,

Pohl,T., Portetelle,D., Puhler,A., Purnelle,B., Ramsperger,U.,

Renard,C., Thebault,P., Vandenbol,M., Weidner,S. and Galibert,F.

Analysis of the chromosome sequence of the legume symbiont

Sinorhizobium meliloti strain 1021

Proc. Natl. Acad. Sci. U.S.A. 98 (17), 9877-9882 (2001)

21396507

11481430

REFERENCE 2 (bases 1 to 273785)

Gouzy,J.

Direct Submission

Submitted (26-JUL-2001) Gouzy J., Submitted on behalf of the MELILO

EU Consortium

MELILO EU Consortium:

Laboratoire de Biologie Molculaire des Relations

Plantes-Microorganismes, UMR215-CNRS-INRA, BP27, F-31326 Castanet,

France, Laboratoire de Genetique et Developpement UMR6061-CNRS,

Faculte de Medecine, 2 avenue du Pr. Leon Bernard, F-35043 Rennes,

France, CMC GmbH, Fritz-Arnold-str. 23, D-78467 Konstanz, Germany,

Universitaet Bielefeld, Biologie IV (Genetik) Universitaetstr 25,

D-33615 Bielefeld, Germany, Unite de Biochimie physiologique,

Universite Catholique de Louvain, Place Croix du Sud 2, Bte 20,

B-1348 Louvain-la-Neuve, Belgium, Unite de Microbiologie, Faculte

des Sciences Agronomiques de Gembloux, Avenue Marechal Juin 6,

B-5030 Gembloux, Belgium. E-mail:Jerome.Gouzy@toulouse.inra.fr

http://sequence.toulouse.inra.fr/meliloti.html.

Location/Qualifiers

1..273785

/organism="Sinorhizobium meliloti"

/strain="1021"

/db_xref="taxon:382"

172..1461

/gene="SMC03265"

172..1461

/gene="SMC03265"

/EC_number="1.4.99.1"

/function="small molecule metabolism; degradation; amino

acids"

/note="Product confidence : putative

Gene name confidence : hypothetical

predicted by Codon_usage

predicted by Homology

predicted by Framed"

/codon_start=1

/evidence=not_experimental

/transl_table=11

/product="PUTATIVE AMINO ACID DEHYDROGENASE TRANSMEMBRANE

PROTEIN"

/protein_id="CAC47696.1"

/db_xref="GI:15076143"

/db_xref="SPTREMBL:Q92LE8"

/translation="MPEVSAIVVAGIVGLSVAYGALNGVSVIMVDRNEGGDKASFG

NAGGTAVTEVPVSPVGLMKVFPWMLDPLGLALRMAHTPLRLPLWLNWFARVGTASE

VRRISTALADINGRVYDDLLPLFNEVGLAGLLQKGLTVEYSENGLRNDHMDWDAKR

DLGVEOELTAEIIRMEPEALSPIVKNGVTPQWSHIKDKPVRVDVLRHLWLVANGABI

VRGEVDVYVTSAGAKVVDGGRMTARKIVIAAGAWSGRKLAKOMGRVILVESRGYT

TTIASPGVALEREIVFADRKVPVPLSIGLRIGGAEEGGLVAPNVAASKALVKLAK

HYFDIREGGVSWGSHRPTTIDSLPVIKRGSPSSNIYAFGHGLGLTQGPITGKIV

SDLLGRKPLIDIAFVSFVARFSLKASFYDSFDLFLR"

1750..2041

/gene="f5 OR SMC03266"

3054..4070

/gene="SMC03267"

3054..4070

/gene="SMC03267"

/function="macromolecule metabolism; macromolecule

degradation; degradation of proteins, peptides,

glycopeptides"

/note="Product confidence : putative

Gene name confidence : hypothetical

predicted by Codon_usage

predicted by Framed"

/codon_start=1

/evidence=not_experimental

/transl_table=11

/product="PUTATIVE DIPEPTIDASE PROTEIN"

/protein_id="CAC47697.1"

/db_xref="GI:15076144"

/db_xref="SPTREMBL:Q92LE7"

/translation="MDAELHGDALFIDAVAPLLHKQVGVSVLGGVTCVAPTGVGV

SESPALKTIGNRLVIRTPQGVLRKARHVEDAKRDKMGILFHFQGTGPFETDL

DLVDPAHAGVRVOLTYNKGVYDGGCEELDGLSRFGKDLVRLNRLIIVDCSH

TYKTTMDALSLTRPAVSHSNVLFPSKRNIRDDOAKAAATGVAVTAPQFV

VSANGAARESIDDLIAHIDHVSFVIGIDQVGLDYWMGTLPFSDLTGVVPEHWHV

EAGDWSQAVPRPHYNPPEIETPELLKGTVGLIRRGYSEEDIRKILGLNWLRYVEH

WVGE"

4070..5947

/gene="SMC03268"

4070..5947

/gene="SMC03268"

/function="cell processes; transport of small molecules;

amino acids, amines, peptides"

/note="Product confidence : putative

Gene name confidence : hypothetical

predicted by Codon_usage

predicted by Homology

predicted by Framed"

/codon_start=1

/evidence=not_experimental

/transl_table=11

/product="PUTATIVE PEPTIDE TRANSPORT SYSTEM ATP-BINDING

ABC TRANSPORTER PROTEIN"

/protein_id="CAC47698.1"

/db_xref="GI:15076145"

/db_xref="SPTREMBL:Q92LE6"

/translation="MARLEFISPLVDVNPVKWESKPVKQVLSVNLNMCSEFRADGLWS

SVYGVSPFDIGARETAVVGVSGSGKSVTALSIMRLNSPRGTIEGSKVLGGKELSL

SDAENRQIRGNDIAMIFQEPMTSLNPLVTIGFQTAESLIILHRLSRAQAEINIRLE

KVRIAPKSRFHEPHFSGGMRQVIMALACKPKLLIADETTDITDITVITQAILL

LLKELODQMSVLFIIDHMGVAEIAIDRVVMYKGOAVETASTDDIFNRPNPTFRS

LLKSVPLKSGMAGRSRPMRFVVDRTGEADVPVEVDIVKVSERPVLEAVGLTKRFE

VKSGLFTVKGCYHVAENVFSIKAGETLALYSGSGCKSTGTGRSVNRLVPEPSGGLVL

LDGVDTVLKLSQNELKHQRNHMQMIFQDPYASLNPRMNIQTALAEPLINKGQOSEVA

DKVAELQKVLSPDMARFHEFSGQQRICIRALALEPKLIVADEAVSALDVSV
KQVNRMLDLQARGLAYLFISHDMVVRVSHVAVMYLGEIVAGSREAIQNPQ
HPYTRRLLSAVPIADPARRLOKRTVSNDEIKSPFPADYVQVPVQRYRTVGHVIMIS
G"

gene
6017. .7615
/gene="SMC03269"
/gene="SMC03269"
/function="cell processes; transport of small molecules;
amino acids, amines, peptides"
/note="Product confidence : putative
Gene name confidence : hypothetical
predicted by Codon_usage
predicted by Homology
predicted by Framed"
/codon_start=1
/evidence=not_experimental
/transl_table=11
/product="PUTATIVE PEPTIDE-BINDING PERIPLASMIC ABC
TRANSPORTER PROTEIN"
/protein_id="CAC47699.1"
/db_xref="GI:15076146"
/db_xref="SPTREMBL:Q92LE5"
/translation="MTPATIKTKTEKIAALKLSTALVAGLLFTSAANADTLRWGML
KIIDPIVTSGLTHIAYMIYDKLFEDESGVPLQAEKVOISADNKSFTIKRLGL
KFSGTPLTSDDVILQSLTRWASRDQTQLLRKTKLDKIDNTLRIDLFESFVPSA
LAGITGNPAFLPKRVASLPTEALTDTGSGPYQDFSTWPKQMRVYTKNSYIIPR
DEASFYAGREAYMDRIEFYVDPDNTAMASMLTKQDIWLDPPMANAVLARNNPL
KVKGYSLOGLFVNHVVPDFNVAKAALTYLIDQKEINAIJVGDPPEFHACYAFLT
CSGTGNEVYDGHGRGPANVEKAKELENEAGYQGPVILDPDTPDEAHARALYLAQQ
LREIGVQVLDQAMGWLITSRRTSKEPADKGNLFPNVMEGQFASPSTHMLASCC
EKANFGPCDKQIEDLRSFIRVVIDEKKKIADELQARASAYLPFVMTGEYNVMSY
GLNIEGFIPSAPEPFVFNRRKQ"

gene
7638. .8582
/gene="SMC03754"
/gene="SMC03754"
/function="cell processes; transport of small molecules;
amino acids, amines, peptides"
/note="Product confidence : putative
Gene name confidence : hypothetical
predicted by Codon_usage
predicted by Homology
predicted by Framed"
/codon_start=1
/evidence=not_experimental
/transl_table=11
/product="PUTATIVE PEPTIDE TRANSPORT SYSTEM PERMEASE ABC
TRANSPORTER PROTEIN"
/protein_id="CAC47700.1"
/db_xref="GI:15076147"
/db_xref="SPTREMBL:Q92LE4"
/translation="MIRYLIRALISAIPVIGVTVIVVGLTVLPGDAASLMAGDYAS
PADIALRVKLHLDDPIITQYIYWLGNLKGDFGHSIFSQQPVLAALILSRVEPTILA
IAATLVAVVAVPLGLLAHNGKGIADQVVSGLGVAGTAIPAFELGYLLIFAFGYIG
SMPIAGFAGTSLLEMAKHVLLPALMGAFATALISRTIRASAEALNQDFVRARA
KGIDETITMHCLEKPIANPVITVGTSTFAHMLGGVITETVFAIPGTGSLVLEAIOH
KDPVITQGVLLLSVYLVNLVLDUSLYLIVDPRIKYT"

gene
8595. .9437
/gene="SMC03272"
/gene="SMC03272"
/function="cell processes; transport of small molecules;
amino acids, amines, peptides"
/note="Product confidence : putative
Gene name confidence : hypothetical
predicted by Codon_usage
predicted by Homology
predicted by Framed"
/codon_start=1
/evidence=not_experimental
/transl_table=11
/product="PUTATIVE PEPTIDE TRANSPORT SYSTEM PERMEASE ABC

TRANSPORTER PROTEIN"

Query Match 76.8%; Score 19.2; DB 1; Length 273785;
Best Local Similarity 87.5%; Pred. No. 4.4e+02;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 CGCGTCGCGTGCGAGCAGCTGACAAA 24
||||||| ||||| || |||||
Db 39090 CGCGTCGCGTGCGAGCAGCTGACAAA 39113

RESULT 3
AP003004 338579 bp DNA linear BCT 15-MAY-2001
Mesorhizobium loti DNA, complete genome, section 11/21.
ACCESSION AP003004 BA000012
VERSION AP003004.2 GI:14024426
KEYWORDS
SOURCE
ORGANISM
Mesorhizobium loti (strain:MAFF303099) DNA.
Mesorhizobium loti
Bacteria: Proteobacteria; alpha subdivision; Rhizobiaceae group;
Phyllobacteriaceae; Mesorhizobium.
REFERENCE
1 (sites)
AUTHORS
Kaneko, T., Nakamura, Y., Sato, S., Asamizu, E., Kato, T., Sasamoto, S.,
Watanabe, A., Idesawa, K., Ishikawa, A., Kawashima, K., Kimura, T.,
Kishida, Y., Kiyokawa, C., Kohara, M., Matsumoto, M., Matsuno, A.,
Mochizuki, Y., Nakayama, S., Nakazaki, N., Shimpo, S., Sugimoto, M.,
Takeuchi, C., Yamada, M. and Tabata, S.
Complete genome structure of the nitrogen-fixing symbiotic
bacterium Mesorhizobium loti
DNA Res. 7 (6), 331-338 (2000)
21082930
2 (bases 1 to 338579)
Direct Submission
AUTHORS
Kaneko, T.
TITLE
Submitted (05-DEC-2000) Takakazu Kaneko, Kazusa DNA Research
Institute, The First Laboratory for Plant Gene Research; Yana
1532-3, Kisarazu, Chiba 292-0812, Japan
(E-mail: kaneko@kazusa.or.jp/
URL: http://www.kazusa.or.jp/rhizobase/
Tel: 81-438-52-3935 (ex.2338), Fax: 81-438-52-3934)
On May 11, 2001 this sequence version replaced gi:11994979.
COMMENT
FEATURES
Location/Qualifiers
1. 338579
/organism="Mesorhizobium loti"
/strain="MAFF303099"
/db_xref="taxon:381"
complement(47. .361)
/gene="ml14353"
complement(47. .361)
/gene="ml14353"
/note="hypothetical protein"
/codon_start=1
/transl_table=11
/protein_id="BAB51030.1"
/db_xref="GI:14024427"
/translation="MRLRHLMAGRNAAGTNPITPTTALVVDGIYRTRNPFFVL
GAFVLVGLGVAAASFVITVLVPLLVINTGVIAPEERYLERKFGDAYRAYRVRRR
WI."
980. .1483
/gene="mlr4354"
980. .1483
/gene="mlr4354"
/note="unknown protein"
/codon_start=1
/transl_table=11
/protein_id="BAB51031.1"
/db_xref="GI:14024428"
/translation="MRGLIALVSSLVLSAAPVLAQAQATKIQHNWAGTYVQASGG
KVCYLVAVPTDKQPTLDHGMDFVFSQRQGVSYEPQFIAGNFQEGSKATVTDK
KSFMSFTRGSAWVENAAEPVLIAAMKTGTDKMKVTAKSGRGNPTSYVFSLSKGISAL
SSIAKCK"
complement(1541. .2884)
gene

CDS	/gene="ml14356" complement(1541. .2884) /gene="ml14356" /codon_start=1 /transl_table=11 /product="chloride channel protein" /protein_id="BAB51032.1" /db_xref="GI:14024429" /translation="MAGTSRKYPMRLRRSRAMWGRVWQPLVFWAGASIGVISVLV AVLADQALFHMIGDGGRRRYPLAITPLGVLCARLAFPPQSGSGIPQAI ARHLRDDRRSHLLSURLAGKALTITVGLFCGASIGREGTVQVGASLMQARWGG MAQARGILAGSAGIAAFAADPLAGIVFAIEEMGRTEARTNGLVTLTAVILAGLASL GLLGNYYTGVSKETIAFAADPMLVIGCGIIGGFGALFSLALKITRRIRRHHTQOP LARALLVAACGLGVAVTIGASGGLTFTGTGAQARGAVEGTPLPWFPEAKFLAGLLS MVSGLPGGIFAPSLAVAGAGISSGLLFLFASAGAAALIGMAGYFAGVVQAPMTAFVII LEMTGNHNVIALMASLGYGTARTMSHEPLYHALSRVFAEAI RRRRAETGVV" complement(3069. .3416) /gene="ml14358" complement(3069. .3416) /gene="ml14358" /note="unknown protein" /codon_start=1 /transl_table=11 /protein_id="BAB51033.1" /db_xref="GI:14024430" /translation="MGQPCLYKMGVCRMNSTKFAVIAGLVGVGLSASQAFEDAMGT MTMKGGVEYMAIMPQGHGTMMPDAKGAEMMKMAKPIKHCMMMTDAKGAFLMIDTS TKKAQAECEKMAAM" 3640. .4872 /gene="mlr4359" 3640. .4872 /gene="mlr4359" /note="hypothetical protein" /codon_start=1 /transl_table=11 /protein_id="BAB51034.1" /db_xref="GI:14024431" /translation="MTLSFDLTTGCARDALARAAPKEKPSLIGLTRAEALVASGI VPERAKMAQOLWHMVMVGVGSDRAGFENISKDLRAELDKHFTVARPEIVEEQISSD GTRKWLFRPPGARGPVEIETVYIPEEGRGTLCISSQVGTCTLCSCFCHTQKLVNR LTTEILAOLLTARDLGRDPDPTPDGAIVPAERGGKGVNVMGMGEPLNDFEAVKK ALLIASDGLSLSKRRITLSTGVVPEIFRTGEGVMLAISLHATDDURLDLVPI NKYPLKELIAACRAYGPLSNKRITFEYVMLKQVNDSDIAKGLIKLKGIPAKINL IPFPWPQNYOCSDWETIEKPTADYINNAGYASPIRTPRGRDILAACGQLKSESERMR KVDRLEAMMITAGHEA" 4677. .5318 /gene="mlr4361" 4677. .5318 /gene="mlr4361" /note="hypothetical protein" /codon_start=1 /transl_table=11 /protein_id="BAB51035.1" /db_xref="GI:14024432" /translation="MLGLGDDEIRLHQOCLADPHTRPRRLRPRAEIGIRA HAKSGPAGAGSHDRGAGGVSRVAGVKRIAIIVAGYIAGVLTSLVPLVLTLLLEG IEYGNWSLVFTFDWGLSPLAVLTAVACALLIVGPAILVAEWLSLRGWFYFVATGAIT SIAGVLVPGSAGGFVFNPGFLGLTATAGTVAGSVWLLAGRRSGRLMPDRPG" 5315. .5797 /gene="mlr4362" 5315. .5797 /gene="mlr4362" /note="hypothetical protein" /codon_start=1 /transl_table=11 /protein_id="BAB51036.1" /db_xref="GI:14024433" /translation="MTRLVAYLVRLAMILFGYAVASLAASAFNLVFLASLGYTPDHT HPTATSLYSPFVAFVAFYAFPAVIVLVAELIGRRDRLYALAGAVATVFLG LVDHAQESTFAITDTSALMAVIGGVMGVIYFVLSAGRWAGSWWKDEASTSPGSGS " complement(5769. .6410) /gene="ml14363"	CDS	/gene="ml14356" complement(1541. .2884) /gene="ml14356" /codon_start=1 /transl_table=11 /product="chloride channel protein" /protein_id="BAB51032.1" /db_xref="GI:14024429" /translation="MAGTSRKYPMRLRRSRAMWGRVWQPLVFWAGASIGVISVLV AVLADQALFHMIGDGGRRRYPLAITPLGVLCARLAFPPQSGSGIPQAI ARHLRDDRRSHLLSURLAGKALTITVGLFCGASIGREGTVQVGASLMQARWGG MAQARGILAGSAGIAAFAADPLAGIVFAIEEMGRTEARTNGLVTLTAVILAGLASL GLLGNYYTGVSKETIAFAADPMLVIGCGIIGGFGALFSLALKITRRIRRHHTQOP LARALLVAACGLGVAVTIGASGGLTFTGTGAQARGAVEGTPLPWFPEAKFLAGLLS MVSGLPGGIFAPSLAVAGAGISSGLLFLFASAGAAALIGMAGYFAGVVQAPMTAFVII LEMTGNHNVIALMASLGYGTARTMSHEPLYHALSRVFAEAI RRRRAETGVV" complement(3069. .3416) /gene="ml14358" complement(3069. .3416) /gene="ml14358" /note="unknown protein" /codon_start=1 /transl_table=11 /protein_id="BAB51033.1" /db_xref="GI:14024430" /translation="MGQPCLYKMGVCRMNSTKFAVIAGLVGVGLSASQAFEDAMGT MTMKGGVEYMAIMPQGHGTMMPDAKGAEMMKMAKPIKHCMMMTDAKGAFLMIDTS TKKAQAECEKMAAM" 3640. .4872 /gene="mlr4359" 3640. .4872 /gene="mlr4359" /note="hypothetical protein" /codon_start=1 /transl_table=11 /protein_id="BAB51034.1" /db_xref="GI:14024431" /translation="MTLSFDLTTGCARDALARAAPKEKPSLIGLTRAEALVASGI VPERAKMAQOLWHMVMVGVGSDRAGFENISKDLRAELDKHFTVARPEIVEEQISSD GTRKWLFRPPGARGPVEIETVYIPEEGRGTLCISSQVGTCTLCSCFCHTQKLVNR LTTEILAOLLTARDLGRDPDPTPDGAIVPAERGGKGVNVMGMGEPLNDFEAVKK ALLIASDGLSLSKRRITLSTGVVPEIFRTGEGVMLAISLHATDDURLDLVPI NKYPLKELIAACRAYGPLSNKRITFEYVMLKQVNDSDIAKGLIKLKGIPAKINL IPFPWPQNYOCSDWETIEKPTADYINNAGYASPIRTPRGRDILAACGQLKSESERMR KVDRLEAMMITAGHEA" 4677. .5318 /gene="mlr4361" 4677. .5318 /gene="mlr4361" /note="hypothetical protein" /codon_start=1 /transl_table=11 /protein_id="BAB51035.1" /db_xref="GI:14024432" /translation="MLGLGDDEIRLHQOCLADPHTRPRRLRPRAEIGIRA HAKSGPAGAGSHDRGAGGVSRVAGVKRIAIIVAGYIAGVLTSLVPLVLTLLLEG IEYGNWSLVFTFDWGLSPLAVLTAVACALLIVGPAILVAEWLSLRGWFYFVATGAIT SIAGVLVPGSAGGFVFNPGFLGLTATAGTVAGSVWLLAGRRSGRLMPDRPG" 5315. .5797 /gene="mlr4362" 5315. .5797 /gene="mlr4362" /note="hypothetical protein" /codon_start=1 /transl_table=11 /protein_id="BAB51036.1" /db_xref="GI:14024433" /translation="MTRLVAYLVRLAMILFGYAVASLAASAFNLVFLASLGYTPDHT HPTATSLYSPFVAFVAFYAFPAVIVLVAELIGRRDRLYALAGAVATVFLG LVDHAQESTFAITDTSALMAVIGGVMGVIYFVLSAGRWAGSWWKDEASTSPGSGS " complement(5769. .6410) /gene="ml14363"	gene	gene		
CDS	/complement(5769. .6410) /gene="ml14363" /note="hypothetical protein" /codon_start=1 /transl_table=11 /protein_id="BAB51037.1" /db_xref="GI:14024434" /translation="MSFIPDTATLIQFAIATVILAITPGDMTLEVSRTLSQGRATGF ASWAGALCCGTLIHTLVVVVSALIVASPMFAFVLKIFGAGYLVLAWQALAKGSAFS PEKKTGQPSILLSRWAGLGVNLLNPKIILFFMTLPQFVSAHDHPACPKLFFLGLME IVLSIVTAPMWLAAEKFSAAKASPRVTRVVDYLFAGVFSAFALKILTAAQAK" complement(6530. .8746) /gene="ml14364" complement(6530. .8746) /gene="ml14364" /note="hypothetical protein" /codon_start=1 /transl_table=11 /protein_id="BAB51038.1" /db_xref="GI:14024435" /translation="MGCIVARDLSKFSRPAKATQNCPCGHATVAKRLLPARCGCLGL GPFLMRVISCIIVTEHNLWLIVLAALMCVTGCWVTIGLDRARKTVQVMRGWLFETAV AAGSSWCFHFIAMLAYQPCAPITPDPAFTMISLVAMVGTGAGFLGALDKRGLAPE WGGCLVGLAISAMHITGMAYHVAGIVENDAFYVVASVLJAMAFSSVAVGHAMRRPYR WSHYVGILLVLAIVGLHFTMAAAVAVITPLSFITGTNPDIILEMAVAVAYAGLIVAA TGFSYLIIDRGRLESFERLOHLALNDSLUGLANRVSFNDRLDHLDEIRAREHODTMTA VIVIDLRFEINDLRHAAGDQALKI IARELAKLTGDEGFVARLGGDEFAAIKRPKD QNDLGLVLSLEKSLFEPLRIDFETVTGASIVAVYPRDGADRELRLVSNADLAMYEA KNVTRAVCFYESAMDETARRALATDLQADIRGELSHTYQVOTSVOTGATCGYEA LNRWHPVGMIPPAEFIPAEENGSLAIGELWLTACROASWONGHKIAVNLSPV LQFHADLTLVHQLLVETGLSPARLELELTESTIVADKVTTLHLVLRQIKALGVTIAD DFGTGYSLDVLTRSPFDPKIKLDRSMADVSPQAKIIRAVLTIGRSLDIPVLAEG VETHVOLTILQVEGCNEAQGYFLGRPKPIDQIGLISGADTGAPTFEVPVMMRVAIG" 8951. .10174 /gene="mlr4366" 8951. .10174 /gene="mlr4366" /codon_start=1 /transl_table=11 /product="argininosuccinate synthase" /protein_id="BAB51039.1" /db_xref="GI:14024436" /translation="MSKTDVKKVVLAYSGGLDTSIILKWLQTELGAEVVTFADLQO GGELEPARKAEMGIKDIRIDVREFFVADFPMPERANTVYEGYLLGTSIARPLI SKHLVDIARTGADAIAGHATGNDQVRELSAYALNPDIKVIAFRDWSFKSRDIL INFAEQHQIPVAKDKGEAPPFSDVANDLLHSSEKGVLEDWPSEPPFHORTVSPMDA PKVTEIEIEFKDPIALNGKLSPATMLAALNDLRDNGIGLDLVERNFYVLMQSR GVYETPGGAILIVAHRAIESITLDRGAHLKDFEPMRYAELIYNGFWFSPEFLMLQAM IDKSQDEVEGTVRLLYKGNVMVTGRSKKTLSDALVTDEDDRGAYDQKDAAGFIRL NALRLTLAARNKKG" 10349. .10786 /gene="mlr4368" 10349. .10786 /gene="mlr4368" /note="unknown protein" /codon_start=1 /transl_table=11 /protein_id="BAB51040.1" /db_xref="GI:14024437" /translation="MKNISILLGAVVVLSVCQSSQAGAKRDKPKVDNNECIASARTSN LTFKHDMAFPMGVSQERMPALFCORLADGIRTSYSDINNNLQDQPTETIMVWLKGG PKRAAPHSQAPRSPFNFCNSGIDGATQVPASQRCPASGYAHY" 10829. .11314 /gene="mlr4369" 10829. .11314 /gene="mlr4369" /note="unknown protein"	CDS	gene	CDS	gene		
Query Match	76.8%;	Score	19.2;	DB	1;	Length	338579;
Best Local Similarity	87.5%;	Pred. No.	4.4e+02;				
Matches	21;	Conservative	0;	Mismatches	3;	Indels	0;
GY	1	CGCGTCGGTGCAGGACGTGACAAA	24				


```

Query Match      74.4%; Score 18.6; DB 8; Length 704;
Best Local Similarity 84.0%; Pred. No. 1.2e+03;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 CCGGTCGGTGCAGGACGTGACAAAT 25
      ||||| ||||| ||||| ||||| |||||
Db 542 CCGGTCGTGTCAGGATGTAACCAAT 566

RESULT 7
VFU41417      3670 bp      DNA      linear      BCT 24-DEC-1996
DEFINITION    Vibrio furnissii beta-N-acetylhexosaminidase gene, complete cds.
ACCESSION     U41417
VERSION       U41417.1 GI:1698441
KEYWORDS
SOURCE        Vibrio furnissii.
ORGANISM      Bacteria; Proteobacteria; gamma subdivision; Vibrionaceae; Vibrio.
REFERENCE     Keyhani,N.O. and Roseman,S.
              The chitin catabolic cascade in the marine bacterium Vibrio
              furnissii. Molecular cloning, isolation, and characterization of a
              periplasmic beta-N-acetylglucosaminidase
              J. Biol. Chem. 271 (52), 33425-33432 (1996)
JOURNAL
MEDLINE       97125983
PUBMED        8969205
REFERENCE     2 (bases 1 to 3670)
AUTHORS       Keyhani,N.O. and Roseman,S.
TITLE         Direct Submission
JOURNAL       Submitted (27-NOV-1995) Nemat O. Keyhani, Biology, Johns Hopkins
              University, 3400 N. Charles St., Baltimore, MD 21218, USA
FEATURES
source        1..3670
              /organism="Vibrio furnissii"
              /db_xref="taxon:29494"
              845..2680
              /codon_start=1
              /transl_table=11
              /product="beta-N-acetylhexosaminidase"
              /protein_id="AAC44672.1"
              /db_xref="GI:1698442"
              /translation="MNYRIDFAVLSEHPQFCRFLTLHNLSDQLKAWSLHFTIDRYI
              OPDSISHOIHGVSPCSLTPEODVINSNHFYCEFSIKTAPFPHYVTDGKAFAFVQ
              INDVERVRHDLVTFIALASPYRERSEJPDATATLSLLPKPIERLDGFEALTAG
              SQTLOSSCAETAATWIKQELTHLYQWPHDIGSADIVLTNPTIDEGNALLSVDRKP
              IRLEASHGTFVHASATLLQVLPDGDNLPHVIVIKDAPREKIRGMLDCARHEHPL
              ERVKRLINOLAHYKNTFHWHLTDDEGWRIEIKSLPQLTDIGAWRGVDEVLEPOYSLL
              TERKGGFYQETREVIAYAAERGIVTPEIDIPGHSRAAIKALPEWLFDEDDQSYR
              SIOYANDVLSPALPGTYRFLDCVLEVAALPSPFHITGADVPDGVWVNSPKQAL
              MAGEGTDFKELQGLLRYAEKLLKLSKGMVGEAHOHGDKVSKDVIYSWLSQAA
              LNCARGQFVILQPGQFTYLDIADYAPPEPGVDWAGVTPLEARYEYFLVEVPEHPD
              LRKRILGIQCALWELVNNQDRMDYMIYPLRLTAGALGGLDTKIPA"
BASE COUNT    858 a 1046 c 958 g 808 t
ORIGIN
Query Match      74.4%; Score 18.6; DB 1; Length 3670;
Best Local Similarity 84.0%; Pred. No. 1.1e+03;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 CCGGTCGGTGCAGGACGTGACAAAT 25
      ||||| ||||| ||||| ||||| |||||
Db 1033 CACGCCGGAGCAGGACGTGATAAT 1057

RESULT 8
AR022345      3670 bp      DNA      linear      PAT 05-DEC-1998
LOCUS
DEFINITION     Sequence 3 from patent US 5792647.
ACCESSION     AR022345
VERSION       AR022345.1 GI:3976407
KEYWORDS
SOURCE        Unknown.

Query Match      74.4%; Score 18.6; DB 6; Length 3670;
Best Local Similarity 84.0%; Pred. No. 1.1e+03;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 CCGGTCGGTGCAGGACGTGACAAAT 25
      ||||| ||||| ||||| ||||| |||||
Db 1033 CACGCCGGAGCAGGACGTGATAAT 1057

RESULT 10
AC130788/c     217812 bp      DNA      linear      HTG 14-AUG-2002
LOCUS
DEFINITION     Bos taurus clone RP42-331024, WORKING DRAFT SEQUENCE, 5 unordered
              pieces.
ACCESSION     AC130788
VERSION       AC130788.1 GI:22218458
KEYWORDS      HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE        COW.
ORGANISM      Bos taurus
              Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
              Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidea;
              Bovidae; Bovinae; Bos.
              1 (bases 1 to 217812)
              Akhter,N., Antonellis,A., Ayele,K., Beckstrom-Sternberg,S.M.,
              Benjamin,B., Blakesley,R.W., Bouffard,G.G., Breen,K., Brinkley,C.,
              Brooks,S., Dietrich,N.L., Granite,S., Guan,X., Gupta,J.,
              Haghini,P., Hansen,N., Ho,S.-L., Idol,J.R., Karlins,E., Laric,P.,
              Lee-Lin,S.-O., Legaspi,R., Maduro,Q.L., Maduro,V.B.,
              McQuillies,E.H., Mastello,C., Maskeri,B., Mastrian,S.D.,
              McCloskey,J.C., McDowell,J., Paguirigan,C., Pearson,R.,
              Portnoy,M.E., Prasad,A., Schueler,M.G., Stantripop,S., Thomas,J.W.,

ORGANISM      Unknown.
REFERENCE     Roseman,S., Bassler,B., Keyhani,N.O., Chitlaru,E., Rowe,C. and
              Yu,C.
              Bacterial catabolism of chitin
              Patent: US 5792647-A 3 11-AUG-1998;
              Location/Qualifiers
              1..3670
              /organism="unknown"
              858 a 1046 c 958 g 808 t
ORIGIN
Query Match      74.4%; Score 18.6; DB 6; Length 3670;
Best Local Similarity 84.0%; Pred. No. 1.1e+03;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 CCGGTCGGTGCAGGACGTGACAAAT 25
      ||||| ||||| ||||| ||||| |||||
Db 1033 CACGCCGGAGCAGGACGTGATAAT 1057

RESULT 9
AR086949      3670 bp      DNA      linear      PAT 07-SEP-2000
LOCUS
DEFINITION     Sequence 3 from patent US 5985644.
ACCESSION     AR086949
VERSION       AR086949.1 GI:10013715
KEYWORDS
SOURCE        Unknown.
ORGANISM      Unknown.
REFERENCE     Roseman,S., Bassler,B., Keyhani,N.O., Chitlaru,E. and Yu,C.
              Bacterial catabolism of chitin
              Patent: US 5985644-A 3 16-NOV-1999;
              Location/Qualifiers
              1..3670
              /organism="unknown"
              858 a 1046 c 958 g 808 t
ORIGIN
Query Match      74.4%; Score 18.6; DB 6; Length 3670;
Best Local Similarity 84.0%; Pred. No. 1.1e+03;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 CCGGTCGGTGCAGGACGTGACAAAT 25
      ||||| ||||| ||||| ||||| |||||
Db 1033 CACGCCGGAGCAGGACGTGATAAT 1057

RESULT 10
AC130788/c     217812 bp      DNA      linear      HTG 14-AUG-2002
LOCUS
DEFINITION     Bos taurus clone RP42-331024, WORKING DRAFT SEQUENCE, 5 unordered
              pieces.
ACCESSION     AC130788
VERSION       AC130788.1 GI:22218458
KEYWORDS      HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE        COW.
ORGANISM      Bos taurus
              Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
              Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidea;
              Bovidae; Bovinae; Bos.
              1 (bases 1 to 217812)
              Akhter,N., Antonellis,A., Ayele,K., Beckstrom-Sternberg,S.M.,
              Benjamin,B., Blakesley,R.W., Bouffard,G.G., Breen,K., Brinkley,C.,
              Brooks,S., Dietrich,N.L., Granite,S., Guan,X., Gupta,J.,
              Haghini,P., Hansen,N., Ho,S.-L., Idol,J.R., Karlins,E., Laric,P.,
              Lee-Lin,S.-O., Legaspi,R., Maduro,Q.L., Maduro,V.B.,
              McQuillies,E.H., Mastello,C., Maskeri,B., Mastrian,S.D.,
              McCloskey,J.C., McDowell,J., Paguirigan,C., Pearson,R.,
              Portnoy,M.E., Prasad,A., Schueler,M.G., Stantripop,S., Thomas,J.W.,

```

Thomas,P.J., Touchman,J.W., Tsurgeon,C., Vogt,J.L., Walker,M.A.,
Wetherby,K.D., Wiggins,L., Young,A., Zhang,L.-H. and Green,E.D.
NISC Comparative Sequencing Initiative
Unpublished
2 (bases 1 to 217812)
Green,E.D.

Direct Submission

Submitted (14-AUG-2002) NIH Intramural Sequencing Center, 8717
Government Circle, Gaithersburg, MD 20877, USA
----- Genome Center

Center: NIH Intramural Sequencing Center

Center code: NISC

Web site: <http://www.nisc.nih.gov>

Contact: nisc_zoo@nhgri.nih.gov

----- Project Information

Center project name: crz

Center clone name: 331024

----- Summary Statistics

Sequencing vector: plasmid; n/a; 100% of reads

Chemistry: Dye-terminator Big Dye; 100% of reads

Assembly program: Phrap; version 0.990319

Consensus quality: 214289 bases at least Q40

Consensus quality: 215318 bases at least Q30

Consensus quality: 215867 bases at least Q20

Insert size: 208000; agarose-1p

Insert size: 217412; sum-of-contigs

Quality coverage: 8.46x in Q20 bases; agarose-1p

Quality coverage: 8.10x in Q20 bases; sum-of-contigs

* NOTE: This is a 'working draft' sequence. It currently

* consists of 5 contigs. The true order of the pieces

* is not known and their order in this sequence record is

* arbitrary. Gaps between the contigs are represented as

* runs of N, but the exact sizes of the gaps are unknown.

* This record will be updated with the finished sequence

* as soon as it is available and the accession number will

* be preserved.

* 1 3022: contig of 3022 bp in length

* 3023 3122: gap of unknown length

* 3123 36890: contig of 33768 bp in length

* 36891 36990: gap of unknown length

* 36991 66601: contig of 29611 bp in length

* 66602 66701: gap of unknown length

* 66702 120837: contig of 54136 bp in length

* 120838 120937: gap of unknown length

* 120938 217812: contig of 96875 bp in length.

Location/Qualifiers

1. .217812

/organism="Bos taurus"

/db_xref="taxon:9913"

/clone="RP42-331024"

/clone_lib="RP42"

1. .3022

/note="assembly_fragment"

3123. .36890

/note="assembly_fragment"

36991. .66601

/note="assembly_fragment"

clone_end:SP6

vector_side:right"

66702. .120837

/note="assembly_fragment"

clone_end:r7

vector_side:left"

120938. .217812

/note="assembly_fragment"

51110 a 58399 c 58136 g 49765 t 402 others

Query Match 74.4%; Score 18.6; DB 2; Length 217812;

Best Local Similarity 84.0%; Pred. No. 8.3e-02;

Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

BASE COUNT 51110 a 58399 c 58136 g 49765 t

ORIGIN

Query Match 74.4%; Score 18.6; DB 2; Length 217812;

Best Local Similarity 84.0%; Pred. No. 8.3e-02;

Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

BASE COUNT 51110 a 58399 c 58136 g 49765 t

ORIGIN

Query Match 74.4%; Score 18.6; DB 2; Length 217812;

Best Local Similarity 84.0%; Pred. No. 8.3e-02;

Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 CGCGTCGGTCGACGACCTGACAAAT 25

||||| ||||| || |||||

Db 150195 CGCGTCGTGTCACACCTGACAAAT 150171

RESULT 11

AP003914/c

LOCUS

DEFINITION

AP003914

Accession

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

FEATURES

Source

BASE COUNT

ORIGIN

Query Match

Best Local Similarity

Matches

Conservative

Mismatches

Indels

Gaps

QY 1 CGCGTCGGTCGACGACCTGGA 20

||||| ||||| |||||

Db 117065 CGCGACGTCGACGACCTGA 117046

RESULT 12

AP004632

LOCUS

DEFINITION

AP004632

Accession

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

FEATURES

Source

BASE COUNT

ORIGIN

Query Match

Best Local Similarity

Matches

Conservative

Mismatches

Indels

Gaps

QY 1 CGCGTCGGTCGACGACCTGA 20

||||| ||||| |||||

Db 117065 CGCGACGTCGACGACCTGA 117046

RESULT 12

AP004632

LOCUS

DEFINITION

AP004632

Accession

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

FEATURES

Source

BASE COUNT

ORIGIN

Query Match

Best Local Similarity

Matches

Conservative

Mismatches

Indels

Gaps

QY 1 CGCGTCGGTCGACGACCTGA 20

||||| ||||| |||||

Db 117065 CGCGACGTCGACGACCTGA 117046

RESULT 12

AP004632

LOCUS

DEFINITION

AP004632

Accession

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

FEATURES

Source

BASE COUNT

ORIGIN

Query Match

Best Local Similarity

Matches

Conservative

Mismatches

Indels

Gaps

QY 1 CGCGTCGGTCGACGACCTGA 20

||||| ||||| |||||

Db 117065 CGCGACGTCGACGACCTGA 117046

RESULT 12

AP004632

LOCUS

DEFINITION

AP004632

Accession

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

FEATURES

Source

BASE COUNT

ORIGIN

Query Match

Best Local Similarity

Matches

Conservative

Mismatches

Indels

Gaps

QY 1 CGCGTCGGTCGACGACCTGA 20

||||| ||||| |||||

Db 117065 CGCGACGTCGACGACCTGA 117046

RESULT 12

AP004632

ORGANISM Oryza sativa (japonica cultivar-group)
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Ehrhartoideae; Oryzaceae; Oryza.

REFERENCE 1
AUTHORS Sasaki, T., Matsumoto, T. and Yamamoto, K.
TITLE Oryza sativa nipponbare(GA3) genomic DNA, chromosome 8, PAC
clone:P0623F08
JOURNAL Published Only in Database (2002)
REFERENCE 2 (bases 1 to 154255)
AUTHORS Sasaki, T., Matsumoto, T. and Yamamoto, K.
TITLE Direct Submission
JOURNAL Submitted (16-JAN-2002) Takuji Sasaki, National Institute of
Agrobiological Sciences, Rice Genome Research Program; Kamondai
2-1-2, Tsukuba, Ibaraki 305-8602, Japan
(E-mail:tsasaki@ias.affrc.go.jp, URL:http://rgp.dna.affrc.go.jp/,
Tel:81-298-38-7441, Fax:81-298-38-7468)

COMMENT
NOTE: It currently consists of 1 contigs. Gaps between the contigs
are represented as runs of N. The order of the pieces is believed
to be correct as given, however the sizes of the gaps between them
are based on estimates that have provided by the submitter. This
sequence will be replaced by the finished sequence as soon as it is
available and the accession number will be preserved.
* NOTE: This is a 'working draft' sequence.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.

FEATURES
source
Location/Qualifiers
1..154255
/organism="Oryza sativa (japonica cultivar-group)"
/cultivar="Nipponbare"
/db_xref="taxon:39947"
/chromosome="8"
/clone="P0623F08"

BASE COUNT 44972 a 32563 c 31861 g 44859 t

ORIGIN

Query Match 73.6%; Score 18.4; DB 2; Length 154255;
Best Local Similarity 95.0%; Pred. No. 1e+03;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CGCGTCGGTGACGACGTGA 20
|||||
Db 115307 CGCGAGGTGACGACGTGA 115326

RESULT 13
LOCUS A46287
DEFINITION Sequence 2 from Patent WO9521927.
ACCESSION A46287
VERSION A46287.1 GI:2300513
KEYWORDS
SOURCE unidentified.
ORGANISM unidentified.
REFERENCE 1 (bases 1 to 24)
AUTHORS Ratcliffe, P.J., Firth, J.D., Harris, A.L. and Pugh, C.W.
TITLE TARGETING GENE THERAPY
JOURNAL Patent: WO 9521927-A 2 17-AUG-1995;
ISIS INNOVATION LIMITED (GB)
FEATURES
source
Location/Qualifiers
1..24
/organism="unidentified"
/db_xref="taxon:32644"

BASE COUNT 6 a 6 c 8 g 4 t

ORIGIN

Query Match 72.0%; Score 18; DB 6; Length 24;
Best Local Similarity 100.0%; Pred. No. 2.7e+03;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 GTGCAGGACGTGACAAAT 25
|||||

Db 8 GTGCAGGACGTGACAAAT 25
|||||

Search completed: January 3, 2003, 23:53:01
Job time : 316.158 secs

Db 7 GTGCAGGACGTGACAAAT 24
|||||

RESULT 14
LOCUS AX023673
DEFINITION Sequence 15 from Patent WO0017371.
ACCESSION AX023673
VERSION AX023673.1 GI:10184034
KEYWORDS
SOURCE Mus sp.
ORGANISM Mus sp.
REFERENCE 1 (bases 1 to 24)
AUTHORS Binley, K.M. and Naylor, S.
TITLE Polynucleotide constructs and uses thereof
JOURNAL Patent: WO 0017371-A 15 30-MAR-2000;
BINLEY KATIE MARY (GB) ; NAYLOR STUART (GB) ; OXFORD BIOMEDICA LTD
(GB)
FEATURES
source
Location/Qualifiers
1..24
/organism="Mus sp."
/db_xref="taxon:10095"

BASE COUNT 6 a 6 c 8 g 4 t

ORIGIN

Query Match 72.0%; Score 18; DB 6; Length 24;
Best Local Similarity 100.0%; Pred. No. 2.7e+03;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 GTGCAGGACGTGACAAAT 25
|||||

Db 7 GTGCAGGACGTGACAAAT 24
|||||

RESULT 15
LOCUS AX048713
DEFINITION Sequence 13 from Patent WO0069908.
ACCESSION AX048713
VERSION AX048713.1 GI:12225858
KEYWORDS
SOURCE Mus sp.
ORGANISM Mus sp.
REFERENCE 1 (bases 1 to 24)
AUTHORS Ratcliffe, P.J., Maxwell, P.H. and Pugh, C.W.
TITLE Interaction between the vhl tumour suppressor and hypoxia inducible
factor, and assay methods relating thereto
JOURNAL Patent: WO 0069908-A 13 23-NOV-2000;
ISIS INNOVATION LIMITED (GB)
FEATURES
source
Location/Qualifiers
1..24
/organism="Mus sp."
/db_xref="taxon:10095"

BASE COUNT 6 a 6 c 8 g 4 t

ORIGIN

Query Match 72.0%; Score 18; DB 6; Length 24;
Best Local Similarity 100.0%; Pred. No. 2.7e+03;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 GTGCAGGACGTGACAAAT 25
|||||

Db 7 GTGCAGGACGTGACAAAT 24
|||||

Search completed: January 3, 2003, 23:53:01
Job time : 316.158 secs

GenCore version 5.1.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 3, 2003, 22:52:02 ; Search time 41.7981 Seconds
(without alignments)
1346.950 Million cell updates/sec

Title: US-09-787-562-1

Perfect score: 25

Sequence: 1 cgcgcgtgcaggacgtgacaaat 25

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 2185239 seqs, 1125999159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N_Geneseq_101002.*

1:	/SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT.*
2:	/SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT.*
3:	/SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT.*
4:	/SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT.*
5:	/SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT.*
6:	/SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT.*
7:	/SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT.*
8:	/SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT.*
9:	/SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT.*
10:	/SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT.*
11:	/SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT.*
12:	/SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT.*
13:	/SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT.*
14:	/SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT.*
15:	/SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT.*
16:	/SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT.*
17:	/SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT.*
18:	/SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT.*
19:	/SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT.*
20:	/SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT.*
21:	/SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT.*
22:	/SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT.*
23:	/SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT.*
24:	/SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	25	100.0	25	21	AAAL1993
2	19	76.0	19	21	AAAL1994
3	18.6	74.4	3670	17	AAAT36388
4	18.6	74.4	3670	20	AAV83129
5	18.6	74.4	3670	21	AAZ38241
6	18	72.0	24	16	AAQ9458
7	18	72.0	24	20	AAZ11422
8	18	72.0	24	21	AAAL2007
9	18	72.0	24	22	AAC88980
					Murine PKG HRE P42
					Murine PKG HRE tru
					Periplasmic Beta-N
					Hexosaminidase enz
					Vibrio furnissii e
					Hypoxia-inducible
					Hypoxia responsive
					Murine HRE mpkg DN
					Murine hypoxic res

10	18	72.0	513	22	AAH20729
11	18	72.0	4768	22	AAH20729
12	18	72.0	4768	22	AAH20729
13	18	72.0	4768	24	ABU42019
14	18	72.0	4768	24	ABU42019
15	18	72.0	4768	24	ABU42019
16	18	72.0	4768	24	ABU42019
17	18	72.0	4768	24	ABU42019
18	18	72.0	4768	24	ABU42019
19	18	72.0	4768	24	ABU42019
20	18	72.0	4768	24	ABU42019
21	18	72.0	4768	24	ABU42019
22	18	72.0	4768	24	ABU42019
23	18	72.0	4768	24	ABU42019
24	18	72.0	4768	24	ABU42019
25	18	72.0	4768	24	ABU42019
26	18	72.0	4768	24	ABU42019
27	18	72.0	4768	24	ABU42019
28	18	72.0	4768	24	ABU42019
29	18	72.0	4768	24	ABU42019
30	18	72.0	4768	24	ABU42019
31	18	72.0	4768	24	ABU42019
32	18	72.0	4768	24	ABU42019
33	18	72.0	4768	24	ABU42019
34	18	72.0	4768	24	ABU42019
35	18	72.0	4768	24	ABU42019
36	18	72.0	4768	24	ABU42019
37	18	72.0	4768	24	ABU42019
38	18	72.0	4768	24	ABU42019
39	18	72.0	4768	24	ABU42019
40	18	72.0	4768	24	ABU42019
41	18	72.0	4768	24	ABU42019
42	18	72.0	4768	24	ABU42019
43	18	72.0	4768	24	ABU42019
44	18	72.0	4768	24	ABU42019
45	18	72.0	4768	24	ABU42019

ALIGNMENTS

RESULT 1

AAAL1993

ID AAAL1993 standard; DNA; 25 BP.

XX AAAL1993;

AC AAAL1993;

DT 14-AUG-2000 (first entry)

XX Murine PKG HRE P42 DNA sequence.

DE HRE; hypoxia response element; hypoxia-inducible factor; HIF; vasotropic;

DE HRE; hypoxia response element; hypoxia-inducible factor; HIF; vasotropic;

DE HRE; hypoxia response element; hypoxia-inducible factor; HIF; vasotropic;

DE HRE; hypoxia response element; hypoxia-inducible factor; HIF; vasotropic;

DE HRE; hypoxia response element; hypoxia-inducible factor; HIF; vasotropic;

DE HRE; hypoxia response element; hypoxia-inducible factor; HIF; vasotropic;

DE HRE; hypoxia response element; hypoxia-inducible factor; HIF; vasotropic;

DE HRE; hypoxia response element; hypoxia-inducible factor; HIF; vasotropic;

DE HRE; hypoxia response element; hypoxia-inducible factor; HIF; vasotropic;

DE HRE; hypoxia response element; hypoxia-inducible factor; HIF; vasotropic;

DE HRE; hypoxia response element; hypoxia-inducible factor; HIF; vasotropic;

DE HRE; hypoxia response element; hypoxia-inducible factor; HIF; vasotropic;

DE HRE; hypoxia response element; hypoxia-inducible factor; HIF; vasotropic;

DE HRE; hypoxia response element; hypoxia-inducible factor; HIF; vasotropic;

DE HRE; hypoxia response element; hypoxia-inducible factor; HIF; vasotropic;

DE HRE; hypoxia response element; hypoxia-inducible factor; HIF; vasotropic;

DE HRE; hypoxia response element; hypoxia-inducible factor; HIF; vasotropic;

DE HRE; hypoxia response element; hypoxia-inducible factor; HIF; vasotropic;

DE HRE; hypoxia response element; hypoxia-inducible factor; HIF; vasotropic;

DE HRE; hypoxia response element; hypoxia-inducible factor; HIF; vasotropic;

DE HRE; hypoxia response element; hypoxia-inducible factor; HIF; vasotropic;

DE HRE; hypoxia response element; hypoxia-inducible factor; HIF; vasotropic;

DE HRE; hypoxia response element; hypoxia-inducible factor; HIF; vasotropic;

DE HRE; hypoxia response element; hypoxia-inducible factor; HIF; vasotropic;

DE HRE; hypoxia response element; hypoxia-inducible factor; HIF; vasotropic;

DE HRE; hypoxia response element; hypoxia-inducible factor; HIF; vasotropic;

DE HRE; hypoxia response element; hypoxia-inducible factor; HIF; vasotropic;

DE HRE; hypoxia response element; hypoxia-inducible factor; HIF; vasotropic;

DE HRE; hypoxia response element; hypoxia-inducible factor; HIF; vasotropic;

```

XX Novel polynucleotide constructs comprising at least two repeats of a
PT hypoxia response element useful for driving expression of nucleic acids
PT of interest in a cell under hypoxic conditions
XX
XX Disclosure; Page 11; 155pp; English.
XX
XX This invention describes novel polynucleotide comprising at least 2
CC repeats of a hypoxia response element (HRE), where the hypoxia-inducible
CC factor (HIF) consensus binding sites within each of the 2 repeats are
CC separated by a spacer of at least 20 contiguous nucleotides. The products
CC of the invention have vasotropic, cardiant, cytostatic and antiarthritic
CC activity and can be used for gene therapy. The polynucleotides are useful
CC for delivering nucleic acids of interest to mammalian cells. Lentiviral
CC vectors are responsive to hypoxic agents and to agents that mimic
CC hypoxia. This regulation can be harnessed in vitro to enhance the
CC production of the vector and can be used in vivo to regulate gene
CC expression in response to a physiological signal. The vectors have
CC utility in disease, where ischaemia, including hypoxia, is a feature,
CC e.g. cardiovascular disease, peripheral arterial disease, cancer and
CC arthritis. The novel regulatory construct is capable of driving very high
CC levels of transcription under conditions of hypoxia whilst providing only
CC low basal levels of transcription under normal oxygen conditions. The
CC polynucleotide construct targets cells within a tumor mass that are under
CC conditions of hypoxia without affecting normal surrounding tissue. This
CC polynucleotide construct targets cells within a tumor mass that are under
CC conditions of hypoxia without affecting normal surrounding tissue. This
CC sequence represents a murine phosphoglycerate kinase (PGK) HRE P24 DNA
CC fragment as described in the method of the invention.
XX
XX Sequence 25 BP; 6 A; 6 C; 9 G; 4 T; 0 other;
SQ
Query Match 100.0%; Score 25; DB 21; Length 25;
Best Local Similarity 100.0%; Pred. No. 0.034;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 CGCGTCGGTGCAGGACGTGACAAAT 25
DB 1 CGCGTCGGTGCAGGACGTGACAAAT 25
RESULT 2
AA11994
ID AA11994 standard; DNA; 19 BP.
AC AA11994;
XX
XX 14-AUG-2000 (first entry)
DE
XX Murine PGK HRE truncated P18 DNA sequence.
XX
XX HRE; hypoxia response element; hypoxia-inducible factor; HIF; vasotropic;
KW cardiant; cytostatic; antiarthritic; gene therapy; ischaemia; arthritis;
KW cardiovascular disease; peripheral arterial disease; cancer;
KW phosphoglycerate kinase; PGK; murine; ds.
XX
XX Mus sp.
XX
XX WO200017371-A1.
XX
XX 30-MAR-2000.
XX
XX 22-SEP-1999; 99WO-GB03181.
XX
XX 23-SEP-1998; 98WO-GB02885.
XX
XX 28-JAN-1999; 99GB-0001906.
XX
XX 16-FEB-1999; 99GB-0003538.
XX
XX (OXFO-) OXFORD BIOMEDICA UK LTD.
XX
XX Binley KM, Naylor S;
PI
XX WPI; 2000-283595/24.
XX
XX Novel polynucleotide constructs comprising at least two repeats of a
PT

```

```

PT hypoxia response element useful for driving expression of nucleic acids
PT of interest in a cell under hypoxic conditions
XX
XX Disclosure; Page 11; 155pp; English.
XX
XX This invention describes novel polynucleotide comprising at least 2
CC repeats of a hypoxia response element (HRE), where the hypoxia-inducible
CC factor (HIF) consensus binding sites within each of the 2 repeats are
CC separated by a spacer of at least 20 contiguous nucleotides. The products
CC of the invention have vasotropic, cardiant, cytostatic and antiarthritic
CC activity and can be used for gene therapy. The polynucleotides are useful
CC for delivering nucleic acids of interest to mammalian cells. Lentiviral
CC vectors are responsive to hypoxic agents and to agents that mimic
CC hypoxia. This regulation can be harnessed in vitro to enhance the
CC production of the vector and can be used in vivo to regulate gene
CC expression in response to a physiological signal. The vectors have
CC utility in disease, where ischaemia, including hypoxia, is a feature,
CC e.g. cardiovascular disease, peripheral arterial disease, cancer and
CC arthritis. The novel regulatory construct is capable of driving very high
CC levels of transcription under conditions of hypoxia whilst providing only
CC low basal levels of transcription under normal oxygen conditions. The
CC polynucleotide construct targets cells within a tumor mass that are under
CC conditions of hypoxia without affecting normal surrounding tissue. This
CC sequence represents a murine phosphoglycerate kinase (PGK) HRE truncated
CC P18 DNA fragment as described in the method of the invention.
XX
XX Sequence 19 BP; 4 A; 4 C; 8 G; 3 T; 0 other;
SQ
Query Match 76.0%; Score 19; DB 21; Length 19;
Best Local Similarity 100.0%; Pred. No. 20;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 4 GTCGGTGCAGGACGTGACA 22
DB 1 GTCGGTGCAGGACGTGACA 19
RESULT 3
AAT36388
ID AAT36388 standard; DNA; 3670 BP.
XX
XX AAT36388;
AC
XX
XX 14-JAN-1997 (first entry)
DE
XX Periplasmic Beta-N-acetylglucosaminidase coding sequence.
XX
XX Periplasmic chitodextrinase; periplasmic Beta-N-acetylglucosaminidase;
KW Beta-N-acetylglucosaminidase; chitin; oligosaccharide; catabolic;
KW catabolism; ss.
XX
XX Vibrio furnissii.
XX
XX Key Location/Qualifiers
FT CDS 845..2680
FT /*tag= a
FT /product= Periplasmic Beta-N-acetylglucosaminidase
XX
XX WO9625424-A1.
XX
XX 22-AUG-1996.
XX
XX 13-FEB-1996; 96WO-US02332.
XX
XX 13-FEB-1995; 95US-0386727.
XX
XX (UJJO ) UNIV JOHNS HOPKINS.
XX
XX Bassler B, Chitlaru E, Keyhani N, Roseman S, Rowe C;
PI Yu C;
XX
XX WPI; 1996-393335/39.
XX
XX P-PSDB; AAW02157.
XX

```


CC compounds. Use of recombinant V. furnissii chitin catabolic enzymes
CC may help to reduce or eliminate these problems.

XX Sequence 3670 BP; 858 A; 1046 C; 958 G; 808 T; 0 other;

XX Query Match 74.4%; Score 18.6; DB 21; Length 3670;

XX Best Local Similarity 84.0%; Pred. No. 47;

XX Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 CGCGTCGTCGAGGACGTGACAAAT 25

DB 1033 CACGCCGAGGACGTCGATAAAT 1057

RESULT 6

AAO99458

ID AAO99458 standard; DNA; 24 BP.

XX AC AAO99458;

XX DT 19-MAR-1996 (first entry)

XX DE Hypoxia-inducible phosphoglycerate kinase-1 expression control sequence.

XX KW Hypoxia; response element; erythropoietin; phosphoglycerate kinase;

XX KW PGK-1; gene therapy; tumour; cancer; P18; P24; ss.

XX OS Mus sp.

XX PN W09521927-A2.

XX PD 17-AUG-1995.

XX PF 15-FEB-1995; 95WO-GB00322.

XX PR 15-FEB-1994; 94GB-0002857.

XX PA (ISIS-) ISIS INNOVATION LTD.

XX PI Firth JD, Harris AL, Pugh CW, Ratcliffe PJ;

XX DR WPI; 1995-293128/38.

XX PT Novel method of targeting gene therapy - using a hypoxically

XX PT inducible expression control sequence linked to a species active

XX PT against disease

XX PS Disclosure; Page 4; 30pp; English.

XX CC AAO99458 and AAO99459 are hypoxia inducible transcription control

XX CC elements P24 and P18, respectively. P24 and P18 may be linked to at

XX CC least one gene encoding a protein that has activity against disease e.g.

XX CC CD2, CD4, etc. Such a construct may be used to treat a patient

XX CC suffering from a disease in which hypoxia is a cause or symptom.

XX CC The constructs are particularly used for treating tumours.

XX CC Sequence 24 BP; 6 A; 6 C; 8 G; 4 T; 0 other;

XX Query Match 72.0%; Score 18; DB 16; Length 24;

XX Best Local Similarity 100.0%; Pred. No. 59;

XX Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 GTGCAGGACGTGACAAAT 25

DB 7 GTGCAGGACGTGACAAAT 24

RESULT 7

AAZ11422

ID AAZ11422 standard; DNA; 24 BP.

XX AC AAZ11422;

XX XX

DT 26-OCT-1999 (first entry)

XX Hypoxia responsive sequence mPKK.

XX Retroviral vector; functional splice donor site; hybrid viral vector;

XX functional splice acceptor site; in vivo gene delivery; therapeutic;

XX lentiviral vector; modified hematopoietic stem cell; MHSC; tumour;

XX ischemia; hypoxia response element; HRE; hypoxia; ss.

XX Mus sp.

XX PN W09915684-A2.

XX PD 01-APR-1999.

XX PF 23-SEP-1998; 98WO-GB02885.

XX PR 25-SEP-1997; 97GB-0020465.

XX PR 23-SEP-1997; 97GB-0020216.

XX PA (OXFO-) OXFORD BIOMEDICA UK LTD.

XX PI Bebbington C, Binley KM, Lewis C, Naylor S;

XX WPI; 1999-263482/22.

XX PT New retroviral vectors, for, e.g. delivering nucleotide sequences to

XX PT solid tumor sites

XX Example 1 (page 68); Fig 1 (page 1/43); 288pp; English.

XX The invention relates to a retroviral vector (RVV) comprising a

XX functional splice donor site (FSDS) and a functional splice acceptor

XX site (FSAS) where: (i) the FSDS and the FSAS flank a first nucleotide

XX sequence of interest (NOI); (ii) the FSDS is upstream of the FSAS; (iii)

XX the RVV is derived from a retroviral pro-vector; (iv) the retroviral

XX pro-vector comprises a first nucleotide sequence (NS) capable of yielding

XX the FSDS and a second NS capable of yielding the FSAS; and (v) the first

XX NS is downstream of the second NS, such that the RVV is formed as a

XX result of reverse transcription of the retroviral pro-vector. A hybrid

XX viral vector (VV) system for in vivo gene delivery, which system

XX comprises a primary VV which encodes a secondary VV, the primary vector

XX capable of infecting a first target cell and of expressing the secondary

XX VV, which secondary vector is capable of transducing a secondary target

XX cell, where the primary vector is obtainable from or is based on a

XX adenoviral vector and the secondary VV is obtainable from or is based on

XX a RVV preferably a lentiviral vector (LVV) is also provided. The systems

XX can be used for delivering NOIs to one or more target sites. The NOIs may

XX encode therapeutic or diagnostic agents. The methods are used

XX particularly for producing modified hematopoietic stem cells (MHSCs) to

XX deliver NOIs to sites such as solid tumours which are characterised by

XX ischemia, such as hypoxia or low glucose concentration. The system

XX permits the stable expression of NOIs in targeted cells, e.g. rapidly

XX dividing cells. Sequences AAZ11420-430 represent nucleotide sequences

XX that are responsive to hypoxia.

XX Sequence 24 BP; 6 A; 6 C; 8 G; 4 T; 0 other;

XX Query Match 72.0%; Score 18; DB 20; Length 24;

XX Best Local Similarity 100.0%; Pred. No. 59;

XX Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 GTGCAGGACGTGACAAAT 25

DB 7 GTGCAGGACGTGACAAAT 24

RESULT 8

AAZ12007

ID AAZ12007 standard; DNA; 24 BP.

XX AC AAZ12007;

XX XX

DT 14-AUG-2000 (first entry)
 DE Murine HRE mPGK DNA.
 KW HRE; hypoxia response element; hypoxia-inducible factor; HIF; vasotropic;
 KW cardiant; cytostatic; antiarthritic; gene therapy; ischaemia; arthritis;
 KW cardiovascular disease; peripheral arterial disease; cancer; murine; ds.
 XX
 OS Mus sp.
 XX WO200017371-A1.
 PN 30-MAR-2000.
 PD 22-SEP-1999; 99WO-GB03181.
 XX 23-SEP-1998; 98WO-GB02885.
 PR 28-JAN-1999; 99GB-0001906.
 PR 16-FEB-1999; 99GB-0003538.
 XX (OXFO-) OXFORD BIOMEDICA UK LTD.
 PA Binley KM, Naylor S;
 XX WPI; 2000-283595/24.
 DR Novel polynucleotide constructs comprising at least two repeats of a
 PT hypoxia response element useful for driving expression of nucleic acids
 PT of interest in a cell under hypoxic conditions
 XX Disclosure; Page 11; 155pp; English.
 XX This invention describes novel polynucleotide comprising at least 2
 CC repeats of a hypoxia response element (HRE), where the hypoxia-inducible
 CC factor (HIF) consensus binding sites within each of the 2 repeats are
 CC separated by a spacer of at least 20 contiguous nucleotides. The products
 CC of the invention have vasotropic, cardiant, cytostatic and antiarthritic
 CC activity and can be used for gene therapy. The polynucleotides are useful
 CC for delivering nucleic acids of interest to mammalian cells. Lentiviral
 CC vectors are responsive to hypoxic agents and to agents that mimic
 CC hypoxia. This regulation can be harnessed in vitro to enhance the
 CC production of the vector and can be used in vivo to regulate gene
 CC expression in response to a physiological signal. The vectors have
 CC utility in disease, where ischaemia, including hypoxia, is a feature,
 CC e.g. cardiovascular disease, peripheral arterial disease, cancer and
 CC arthritis. The novel regulatory construct is capable of driving very high
 CC levels of transcription under conditions of hypoxia whilst providing only
 CC low basal levels of transcription under normal oxygen conditions. The
 CC polynucleotide construct targets cells within a tumor mass that are under
 CC conditions of hypoxia without affecting normal surrounding tissue. This
 CC sequence represents a murine HRE DNA fragment which is described in the
 CC method of the invention.
 XX
 SQ Sequence 24 BP; 6 A; 6 C; 8 G; 4 T; 0 other;
 Query Match 72.0%; Score 18; DB 21; Length 24;
 Best Local Similarity 100.0%; Pred. No. 59;
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 8 GTGCAGGACGTGACAAAT 25
 Db ||||||
 7 GTGCAGGACGTGACAAAT 24
 RESULT 9
 AAC88980
 ID AAC88980 standard; DNA; 24 BP.
 AC AAC88980;
 XX 06-MAR-2001 (first entry)
 DT Murine hypoxic response element PGK-1 HRE.
 XX
 DE

XX Mouse; human; HIF-1alpha; von Hippel-Lindau syndrome protein; VHL;
 KW hypoxia inducible factor-1; cancer; ischaemia; ds.
 XX Mus sp.
 OS WO200069908-A1.
 PN 23-NOV-2000.
 PD 12-MAY-2000; 2000WO-GB01826.
 PF 12-MAY-1999; 99GB-0011047.
 PR (ISTS-) ISIS INNOVATION LTD.
 XX Ratcliffe PJ, Maxwell PH, Pugh CW;
 XX WPI; 2001-025006/03.
 DR Assaying for von Hippel Lindau (VHL)-hypoxia inducible factor (HIF)
 PT alpha subunit interaction modulators for treating ischemia by
 PT contacting a VHL protein and an HIF subunit protein with a putative
 PT modulator
 XX Examples; Page 43; 56pp; English.
 PS The present invention describes a novel assay for use in identifying
 XX modulators of the von Hippel-Lindau protein (VHL) and hypoxia inducible
 CC factor-1 alpha subunit (HIF-1alpha) interaction. The assay comprises
 CC contacting the VHL protein, the HIF-1alpha subunit and the putative
 CC modulator under conditions where the former two would normally complex.
 CC Modulators of this type are useful in the treatment of cancer and
 CC ischaemic conditions such as coronary, cerebral and vascular
 CC insufficiency.
 XX Sequence 24 BP; 6 A; 6 C; 8 G; 4 T; 0 other;
 SQ Query Match 72.0%; Score 18; DB 22; Length 24;
 Best Local Similarity 100.0%; Pred. No. 59;
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 8 GTGCAGGACGTGACAAAT 25
 Db ||||||
 7 GTGCAGGACGTGACAAAT 24
 RESULT 10
 AAH20729
 ID AAH20729 standard; DNA; 513 BP.
 XX AAH20729;
 AC 13-AUG-2001 (first entry)
 DT Murine phosphoglycerate kinase promoter DNA.
 XX Phosphoglycerate kinase; promoter; gene therapy; amniocyte; cytostatic;
 DE E1A region; E1B region; modified tropism; tumor; ds.
 KW Mus sp.
 XX WO200136615-A2.
 PN 25-MAY-2001.
 PD 07-NOV-2000; 2000WO-EP10992.
 XX 18-NOV-1999; 99DE-1055558.
 PR (KOCH/) KOCHANNEK S.
 PA Kochanek S, Schiedner G;
 XX

XX T243 to produce disruption in the DNA. The invention describes
 CC methods of producing embryonic stem (ES) cells comprising a heterozygous
 CC disruption in a target DNA sequence (preferably T243) encoding a TRP and
 CC of producing a knockout mouse comprising a homozygous disruption in a
 CC gene encoding TRP, where the disruption inhibits the production of the
 CC wild type TRP. The invention also relates to identifying agents capable
 CC of affecting a phenotype of a knockout mouse. Also described are methods
 CC of determining whether expansion of the trinucleotide repeat in a gene
 CC encoding TRP produces a phenotypic change. The transgenic animals and
 CC the cells are useful for identifying compounds capable of ameliorating
 CC disease symptoms, and as test substrates for the identification of drugs,
 CC pharmaceuticals, therapies and interventions which may be effective in
 CC treating trinucleotide repeat disorders e.g. fragile X syndrome and
 CC Huntington's disease. The animal models for trinucleotide repeat
 CC disorders are ideal model systems to study the progression of disease in
 CC vivo, the molecular basis of these diseases and show the features
 CC observed in human disease. Using the mice, it is possible to model both
 CC the pathogenic mechanism and the trinucleotide repeat instability in the
 CC mouse.
 SQ Sequence 4768 BP; 1124 A; 1218 C; 1269 G; 1157 T; 0 other;
 Query Match 72.0%; Score 18; DB 22; Length 4768;
 Best Local Similarity 100.0%; Pred. No. 92;
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 8 GTGCAGGACGTGACAAAT 25
 |||||
 Db 2824 GTGCAGGACGTGACAAAT 2841
 RESULT 12
 ABL42019
 ID ABL42019 standard; DNA: 4768 BP.
 XX
 AC ABL42019;
 XX
 DT 11-JUN-2002 (first entry)
 XX
 DE Nucleotide sequence of vector pDG2.
 XX
 KW pDG2; transgenic animal; matrix metalloproteinase-23 gene; MMP-23 gene;
 KW ss.
 XX
 OS Synthetic.
 XX
 PN US2002023275-A1.
 XX
 PD 21-FEB-2002.
 XX
 PF 17-MAY-2001; 2001US-0861077.
 XX
 PR 17-MAY-2000; 2000US-204972P.
 PR 29-JUN-2000; 2000US-215394P.
 XX
 PA (LEVI/) LEVITEN M W.
 XX
 PI Leviten MW;
 XX
 DR WPI; 2002-255684/30.
 XX
 PT Non-human transgenic animal useful as a model for disease and for
 PT identifying agents that modulate gene expression and gene function,
 PT comprises a disruption in the matrix metalloproteinase-23 gene -
 XX
 PS Example 1; Fig 2B; 38pp; English.
 XX
 CC The present sequence represents vector pDG2. This vector contains an
 CC ampicillin resistance gene and a neomycin gene. The vector is used in
 CC the invention. The specification describes a non-human transgenic animal
 CC comprising a disruption in the matrix metalloproteinase (MMP)-23 gene.
 CC Transgenic animals of the invention comprising a homozygous or
 CC heterozygous disruption in MMP23 gene are useful for identifying agents

XX WPI; 2001-343817/36.
 XX
 PT New permanent amniocyte cell lines, useful for producing viral gene
 PT therapy vectors or mutant adenoviruses, express the adenoviral E1A and
 PT E1B gene products -
 XX
 PS Example 1; Page 61; 72pp; German.
 XX
 CC This invention describes novel permanent amniocyte cell lines (A),
 CC containing at least one nucleic acid (I) that causes expression of the
 CC gene products (II) of the adenoviral E1A and E1B regions. (A) are used to
 CC produce gene therapy vectors, especially adeno, adeno-associated, retro
 CC or lentiviral vectors, particularly first- or second generation,
 CC large-capacity or deleted adenoviral vectors. (A) are also used to
 CC produce adenoviral mutants, optionally with modified tropism. The vectors
 CC may express a wide range of therapeutic proteins or antisense RNAs.
 CC Adenoviral mutants, unable to express the E1B 55 kDa protein, are useful
 CC for treating tumors, they replicate in the cells but not significantly in
 CC normal primary cells. (A) can be made efficiently, simply and
 CC reproducibly. The products of the invention have cytostatic activity.
 CC This sequence represents the murine phosphoglycerate kinase promoter
 CC found in plasmid STK146.
 XX
 SQ Sequence 513 BP; 79 A; 171 C; 161 G; 102 T; 0 other;
 Query Match 72.0%; Score 18; DB 22; Length 513;
 Best Local Similarity 100.0%; Pred. No. 76;
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 8 GTGCAGGACGTGACAAAT 25
 |||||
 Db 221 GTGCAGGACGTGACAAAT 238
 RESULT 11
 AAS05243
 ID AAS05243 standard; DNA: 4768 BP.
 XX
 AC AAS05243;
 XX
 DT 07-SEP-2001 (first entry)
 XX
 DE Plasmid vector pDG2 used as a construct for TRP genes.
 XX
 KW Trinucleotide repeat protein; TRP; T243; embryonic stem cell; ES; pDG2;
 KW transgenic animal; knockout mouse; triplet repeat expansion;
 KW fragile X syndrome; Huntington's disease; cyclic; circular; ds.
 XX
 OS Synthetic.
 XX
 PN W0200130798-A1.
 XX
 PD 03-MAY-2001.
 XX
 PF 26-OCT-2000; 2000WO-US29382.
 XX
 PR 26-OCT-1999; 99US-0161488.
 XX
 PA (DELTA-) DELTAGEN INC.
 XX
 PI Klein R, Matthews W, Moore M, Allen KD;
 XX
 DR WPI; 2001-300473/31.
 XX
 PT Novel transgenic animals useful as animal model for characterization of
 PT function of a gene encoding trinucleotide repeat proteins (TRPs),
 PT contains heterozygous disruption in a gene encoding TRP -
 XX
 PS Disclosure; Fig 2B; 106pp; English.
 XX
 CC The present sequence for plasmid vector pDG2 is used as a construct
 CC for genes encoding trinucleotide repeat proteins (TRP) such as gene

CC which modulate MMP23 expression or function. They are also useful for
CC identifying agents that are capable of ameliorating a phenotype of a
CC transgenic animal comprising a disruption in an MMP-23 gene or
CC ameliorating a disease associated with the phenotype of a transgenic
CC animal comprising a disruption in the MMP-23 gene. The animals are
CC useful as an animal model for diseases, disorders and conditions
CC characterized by a disruption in a gene encoding a metalloproteinase,
CC more particularly disease, disorders and conditions associated with the
CC phenotypes demonstrated by the knockout mice. The transgenic animals
CC are useful as test substrates for identification of drugs,
CC pharmaceuticals and therapies effective in treating diseases, disorders
CC and conditions associated with disruption in the target gene. The
CC animal is useful for testing and developing new treatments relating
CC to behavioural phenotypes demonstrated by the animal models.

XX SQ Sequence 4768 BP; 1124 A; 1218 C; 1269 G; 1157 T; 0 other;
Query Match 72.0%; Score 18; DB 24; Length 4768;
Best Local Similarity 100.0%; Pred. No. 92;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 8 GTGCAGGACGTGACAAAT 25
Db 2824 GTGCAGGACGTGACAAAT 2841
|||||

RESULT 13
AAD28659
ID AAD28659 standard; DNA; 4768 BP.

XX AC AAD28659;

XX DT 07-MAY-2002 (first entry)

XX DE Plasmid pDG2 vector.

XX KW Plasmid pDG2; DNA construct; embryonic stem cell; cell disruption; Neo';
XX neomycin; ampicillin resistance gene; ds.

XX OS Unidentified.

XX PN WO200204621-A2.

XX PD 17-JAN-2002.

XX PF 11-JUL-2000; 2000WO-US18812.

XX PR 11-JUL-2000; 2000WO-US18812.

XX PA (DELT-) DELTAGEN INC.

XX PI Klein RD, Brennan TJ;

XX DR WPI; 2002-164642/21.

XX PT Novel nucleotide construct for generating DNA constructs for
XX introducing into embryonic stem cell, comprising a sequence encoding a
XX positive selection marker flanked by restriction enzyme sites -

XX PS Claim 10; Fig 2B; 64pp; English.

XX CC The invention relates to nucleotide construct for generating DNA
XX constructs. The nucleotide construct comprises a sequence encoding a
XX positive selection marker flanked by restriction enzyme sites, where
XX restriction site is flanked by sequences which are not complementary to
XX each other and which do not include at least one type of base at any
XX position, where the construct can be treated so that single-stranded
XX regions are created at each sequence lacking at least one nucleotide.
XX The nucleotide construct is useful in a rapid and efficient method for
XX generating DNA constructs suitable for introduction into embryonic stem
XX cells and for disrupting the function of a gene in a cell. The present
XX sequence is plasmid pDG2 vector construct containing an ampicillin
XX resistance gene and neomycin gene (Neo'). On each site of the Neo' gene

CC are two sites for ligation independent cloning along with restriction
XX sites.

SQ Sequence 4768 BP; 1124 A; 1218 C; 1269 G; 1157 T; 0 other;
Query Match 72.0%; Score 18; DB 24; Length 4768;
Best Local Similarity 100.0%; Pred. No. 92;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 GTGCAGGACGTGACAAAT 25
Db 2824 GTGCAGGACGTGACAAAT 2841
|||||

RESULT 14
AAS17143
ID AAS17143 standard; DNA; 4768 BP.

XX AC AAS17143;

XX DT 14-FEB-2002 (first entry)

XX DE Gene targeting vector pDG2.

XX KW pDG2; ds; retina-specific nuclear receptor; gene targeting;
XX lymphoid-specific GPCR; melanocyte stimulating hormone receptor;
XX magnesium-dependent protein phosphatase; transgenic animal;
XX chemokine receptor 1-like protein; CGMP phosphodiesterase;
XX sulfoltransferase gene; tumour; cancer; retinal degeneration;
XX retinitis pigmentosa.

XX OS Escherichia coli.

XX OS Synthetic.

XX PN WO200167855-A2.

XX PD 20-SEP-2001.

XX PF 16-MAR-2001; 2001WO-US08664.

XX PR 16-MAR-2000; 2000US-190348P.

XX PR 22-MAR-2000; 2000US-191128P.

XX PR 22-MAR-2000; 2000US-191129P.

XX PR 22-MAR-2000; 2000US-191142P.

XX PR 22-MAR-2000; 2000US-191235P.

XX PR 22-MAR-2000; 2000US-191236P.

XX PR 22-MAR-2000; 2000US-191240P.

XX PR 15-MAY-2000; 2000US-204227P.

XX PR 15-MAY-2000; 2000US-204230P.

XX PR 29-JUN-2000; 2000US-215214P.

XX PR 06-JUL-2000; 2000US-216249P.

XX PR 06-JUL-2000; 2000US-216264P.

XX PR 06-JUL-2000; 2000US-216765P.

XX PR 12-JUL-2000; 2000US-218075P.

XX PR 19-JUL-2000; 2000US-219167P.

XX PR 19-JUL-2000; 2000US-219182P.

XX PR 27-JUL-2000; 2000US-221485P.

XX PR 07-AUG-2000; 2000US-223173P.

XX PA (DELT-) DELTAGEN INC.

XX PI Allen KB, Guenther C, Phillips R;

XX DR WPI; 2002-041167/05.

XX PT New targeting construct comprising a first and a second polynucleotide
XX homologous to a target gene, and a selectable marker, useful for
XX introducing targeted mutations into embryonic cells -

XX PS Example 3; Fig 2B; 105pp; English.

XX CC The invention relates to a targeting construct comprising two sequences
XX homologous to a target gene, and a selectable marker, is new. The target

CC gene is a retina-specific nuclear receptor gene, a lymphoid-specific GPCR
CC (G protein coupled receptor) gene, a melanocyte stimulating hormone
CC receptor gene, a magnesium-dependent protein phosphatase gene, chemokine
CC receptor 1-like protein gene, a cGMP phosphodiesterase gene, or a
CC sulfolipase gene. Also included are transgenic mice comprising a
CC disruption in a target gene, where the mouse exhibits an eye abnormality,
CC cellular infiltration, hypoaffective behaviour, lung abnormality, elevated
CC white blood cell count, abnormality in the aorta, kidney, liver, lymph
CC nodes, skin or salivary gland, increased body and organ weight, or
CC elevated levels of ALT (not defined), phosphorus, potassium, or
CC bilirubin, aggressive, hyperactive, increased activity or decreased
CC anxiety behaviour. The construct is used for introducing targeted
CC mutations into embryonic cells. The animal and cell-based systems may be
CC used as models for diseases or conditions associated with physiological,
CC histological or behavioural phenotypes relating to a disruption in a
CC target gene (e.g. tumours, cancer, retinal degeneration and retinitis
CC pigmentosa) and in screening or identifying compounds capable of
CC ameliorating or treating diseases. The present sequence is the vector
CC pDG2 used to generate the gene targeting construct of the invention.
XX
SQ Sequence 4768 BP; 1124 A; 1218 C; 1269 G; 1157 T; 0 other;

Query Match 72.0%; Score 18; DB 24; Length 4768;
Best Local Similarity 100.0%; Pred. No. 92;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 GTGCAGGACGTGACAAAT 25
|||||

Db 2824 GTGCAGGACGTGACAAAT 2841

RESULT 15

AAD09280
ID AAD09280 standard; DNA; 4847 BP.

XX
AC AAD09280;

DT 12-SEP-2001 (first entry)

XX PGK-cre-pA vector DNA.

DE DNA recombinase domain; protein transduction domain; PTD;
XX gene alteration; fusion protein; Human immunodeficiency virus;
KW HIV; pgk-cre-pA vector; ds.

XX Unidentified.

XX WO200149832-A2.

PN 12-JUL-2001.

PD 05-JAN-2001; 2001WO-EP00060.

PF 07-JAN-2000; 2000EP-0100351.

XX 10-NOV-2000; 2000EP-0124595.

XX (ARTE-) ARTEMIS PHARM GMBH.

XX Schwenk F;

PI WPI; 2001-441873/47.

XX
XX Using site-specific DNA recombinase domain/protein transduction domain
PT fusion proteins for inducing target gene alterations in organisms or
PT cell cultures -

XX Example; Page 67-68; 85pp; English.

XX The present invention relates to use of fusion proteins comprising
CC a site-specific DNA recombinase domain e.g. Cre and a protein
CC transduction domain (PTD) e.g. the Human immunodeficiency virus
CC (HIV) derived TAR peptide, for preparing an agent for inducing
CC target gene alterations in a living organism or cell culture. The

XX

CC present invention also provides a method for inducing gene
CC alterations in living organisms using the fusion proteins of the
CC invention. The present sequence is a pgk-cre-pA DNA.

XX
SQ Sequence 4847 BP; 1139 A; 1238 C; 1283 G; 1184 T; 3 other;

Query Match 72.0%; Score 18; DB 22; Length 4847;
Best Local Similarity 100.0%; Pred. No. 92;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 GTGCAGGACGTGACAAAT 25
|||||

Db 2466 GTGCAGGACGTGACAAAT 2483

Search completed: January 3, 2003, 23:03:43
Job time : 44.7981 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 3, 2003, 22:54:17 ; Search time 332.413 Seconds
(without alignments)
1218.024 Million cell updates/sec

Title: US-09-787-562-1

Perfect score: 25

Sequence: 1 cgcgcgtgcagcagtgacaaat 25

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

```
EST:*
1: em_estba:*
2: em_esthum:*
3: em_esthum:*
4: em_esthum:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_hic:*
9: gb_est1:*
10: gb_est2:*
11: gb_hic:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: gb_gss:*
18: em_gss_hum:*
19: em_gss_inv:*
20: em_gss_pln:*
21: em_gss_vrt:*
22: em_gss_fun:*
23: em_gss_mam:*
24: em_gss_mus:*
25: em_gss_other:*
26: em_gss_pro:*
27: em_gss_rod:*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	19.8	79.2	610	12	BG220636
2	19.2	76.8	824	12	BF253450
3	18.6	74.4	541	17	TA249A12Q
4	18.6	74.4	786	12	BG330319
5	17.8	71.2	516	17	AQ235963
6	17.8	71.2	865	17	CNS04REF
					AL294720 Tetraodon

c	7	17.8	71.2	1031	17	CNS04F4W
c	8	17.6	70.4	199	10	BE190622
c	9	17.6	70.4	412	17	AQ647334
c	10	17.6	70.4	414	17	AQ641517
c	11	17.6	70.4	442	17	AZ247056
c	12	17.6	70.4	449	10	BB839535
c	13	17.6	70.4	464	17	AQ844401
c	14	17.6	70.4	515	9	AA957358
c	15	17.6	70.4	564	10	AW664262
c	16	17.6	70.4	577	13	BI443703
c	17	17.6	70.4	578	9	AA386752
c	18	17.6	70.4	592	9	AI981217
c	19	17.6	70.4	615	9	AJ393812
c	20	17.6	70.4	636	17	AG063686
c	21	17.6	70.4	648	10	AV609456
c	22	17.6	70.4	673	14	BQ508828
c	23	17.6	70.4	688	17	BH670819
c	24	17.6	70.4	690	9	AJ447187
c	25	17.6	70.4	714	9	AJ449984
c	26	17.6	70.4	735	9	AJ447188
c	27	17.6	70.4	740	17	BH446078
c	28	17.6	70.4	765	9	AJ456336
c	29	17.6	70.4	766	9	AJ452191
c	30	17.6	70.4	816	10	AW983486
c	31	17.6	70.4	851	17	BH716762
c	32	17.6	70.4	972	12	BG461126
c	33	17.4	69.6	584	10	AW126461
c	34	17.4	69.6	741	13	BM073359
c	35	17.4	69.6	928	13	BI452891
c	36	17.4	69.6	1021	17	CNS05218
c	37	17.4	69.6	1722	11	AY103550
c	38	17.2	68.8	220	13	BJ266315
c	39	17.2	68.8	289	9	AI396788
c	40	17.2	68.8	393	9	AJ331186
c	41	17.2	68.8	437	9	AJ432011
c	42	17.2	68.8	448	10	BB839526
c	43	17.2	68.8	455	9	AI332263
c	44	17.2	68.8	479	9	AA547770
c	45	17.2	68.8	480	9	AJ478558

ALIGNMENTS

RESULT 1	BG220636	610 bp	mRNA	linear	EST 21-APR-2001
LOCUS	RST40423	Athersys	RAGE Library	Homo sapiens	CDNA, mRNA sequence.
DEFINITION	BG220636				
ACCESSION	BG220636.1				
VERSION	EST				
KEYWORDS	human.				
SOURCE	human.				
ORGANISM	Homo sapiens				
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.				
AUTHORS	Harrington, J.J., Sherif, B., Rundlett, S., Jackson, P.D., Perry, R., Cain, S., Leventhal, C., Thornton, M., Ramachandran, R., Whittington, J., Berner, L., Costanzo, D., McElligott, K., Booser, S., Mays, R., Smith, E., Veloso, N., Klika, A., Hess, J., Cothren, K., Lo, K., Offenbacher, J., Danzig, J., and Ducar, M.				
TITLE	Creation of genome-wide protein expression libraries using random activation of gene expression				
JOURNAL	Nat. Biotechnol. 19 (5), 440-445 (2001)				
MEDLINE	21227151				
COMMENT	Contact: Scott J. Cain Athersys, Inc. 3201 Carnegie Ave, Cleveland, OH 44115, USA Tel: 216 431 9900 Fax: 216 361 9596 Email: scain@atersys.com High quality sequence stop: 372. Location/Qualifiers				

Mon Jan 6 15:20:17 2003

```

source
1. .610
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="Athersys RAGE Library"
/cell_line="HT1080"
/notes="See 'Creation of Genome-wide Protein Expression
Libraries using Random Activation of Gene Expression',
Nature Biotechnology, in press. Note that even though the
cell type indicated is HT1080, since a random activation
method was used, these sequence tags are not necessarily
expressed in HT1080 under normal circumstances."
BASE COUNT 113 a 206 c 186 g 104 t 1 others
ORIGIN

Query Match 79.2%; Score 19.8; DB 12; Length 610;
Best Local Similarity 91.3%; Pred. No. 3.3e+02;
Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 CGCGTCGGTGCAGCAGGTGACAA 23
||||| ||||||| |||||||
Db 321 CGCGTCGGTGCAGCAGGTGACAA 343
||||| ||||||| |||||||

RESULT 2
BF253450 824 bp mRNA linear EST 22-OCT-2001
LOCUS HVSMEF0001H07f Hordeum vulgare seedling root EST library HVCDNA0007
DEFINITION (Etiolated and unstressed) Hordeum vulgare cDNA clone
HVSMEF0001H07f, mRNA sequence.
ACCESSION BF253450
VERSION 1
KEYWORDS EST.
SOURCE Hordeum vulgare.
ORGANISM Hordeum vulgare
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooidae
; Triticeae; Hordeum.
1 (bases 1 to 824)
Wing, R., Close, T.J., Klein, R., Rambo, T., Simmons, J., Choi, D.W., Fenton
Y., Henry, D., Palmer, M., Rambo, T., Simmons, J., Choi, D.W., Fenton
R.D., Oates, R. and Main, D.
Development of a genetically and physically anchored EST resource
for barley genomics: Morex unstressed seedling root cDNA library
Unpublished (2001)
On Nov 16, 2000 this sequence version replaced gi:11182651.
Contact: Wing RA
Clemson University Genomics Institute
Clemson University
100 Jordan Hall, Clemson, SC 29634, USA
Tel: 864 656 7288
Fax: 864 656 4293
Email: rwing@clemson.edu
Total hg bases = 118
Seq primer: AATAACCCCTCACTAAAGGG
High quality sequence start: 10
High quality sequence stop: 183.
Location/Qualifiers
1. .824
/organism="Hordeum vulgare"
/cultivar="Morex"
/db_xref="taxon:4513"
/clone="HVSMEF0001H07f"
HVCDNA0007 (Etiolated and unstressed)
/lab_host="TJC121"
/lab_type="Seedling root"
/notes="Vector: lambda2AP; Site_1: EcoRI; Site_2: XhoI;
Seeds were surface sterilized then germinated under axenic
conditions in the dark at room temperature on filter paper
with water, nystatin and cefotaxime in covered
crystallization dishes. Five-day old seedling roots were
then harvested, total RNA was prepared, poly(A) RNA was
purified, one primary unamplified cDNA library was made,

```

and 1 million pfu were in vivo excised to give pBluescript SK(-) cDNA phagemids. These steps were performed in the TJ Close laboratory at the University of California, Riverside (Choi, Close, Fenton). Phagemids were plated and picked at the Clemson University Genomics Institute (CUGI) (Begum, Palmer, Frisch, Atkins and Wing). Plasmid DNA preparations, DNA sequencing and sequence analysis were performed at CUGI (Wing, Yu, Frisch, Henry, Simmons, Oates, Rambo, Main). The sequence has been trimmed to remove vector sequence and contains a minimum of 100 bases of phred value 20 or above. For more details on library preparation and sequence analysis see <http://www.genome.clemson.edu/projects/barley>. To order this clone see <http://www.genome.clemson.edu/orders> see Close TJ, Wing R, Klein, R., Rambo, T., Simmons, J., Choi, D.W., Fenton Y., Henry, D., Palmer, M., Rambo, T., Simmons, J., Choi, D.W., Fenton R.D., Oates, R. and Main, D. Genetically and physically anchored EST resources for barley genomics. Barley Genetics Newsletter 31:29-30. (<http://wheat.pw.usda.gov/ggpages/bgn/31/cover.html>)

```

BASE COUNT 163 a 249 c 254 g 158 t
ORIGIN

```

```

Query Match 76.8%; Score 19.2; DB 12; Length 824;
Best Local Similarity 87.5%; Pred. No. 6.1e+02;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

```

```

QY 1 CGCGTCGGTGCAGCAGGTGACAAA 24
||||| ||||||| |||||||
Db 201 CGCGTCGGTGCAGCAGGTGACAAA 224
||||| ||||||| |||||||

```

```

RESULT 3
TA249A120/c 541 bp DNA linear GSS 13-DEC-2000
LOCUS T. brucei sheared genomic DNA clone 249a12, reverse sequence,
DEFINITION genomic survey sequence.
ACCESSION AL482071
VERSION 1
KEYWORDS GSS.
SOURCE Trypanosoma brucei.
ORGANISM Trypanosoma brucei
Eukaryota; Euzlenozoa; Kinetoplastida; Trypanosomatidae;
Trypanosoma.
1 (bases 1 to 541)
Hall, N., Bowman, S., Lennard, N.J., Doggett, J., Atkin, R.,
Chillingworth, C., Ormond, D., Harris, B., El-Sayed, N., Hou, L.,
Melville, S.E., Rajadream, M.A. and Barrell, B.G.
Direct Submission
Submitted (10-DEC-2000) Trypanosoma brucei genome sequencing
project, Sanger Centre, The Wellcome Trust Genome Campus, Hinxton,
Cambridge CB10 1SA, E-mail: barrel@sanger.ac.uk and
nh@sanger.ac.uk
Constructed at the Institute for Genomic Research (TIGR),
Rockville, MD. Genomic DNA isolated from a cloned population of
Trypanosoma brucei (TREU927/4 GUTAT 10.1) was mechanically sheared
to give a tight size distribution (
4 kb). The v + i method used for the library construction is
described in detail in Smith, H. and Venter, J.C. (Making small
insert libraries for whole genome shotgun sequencing projects. In
Genome Sequencing: A Practical Approach, eds. M. Vaudin and B.
Barrell, Oxford University Press, 1999).
Email: nelsayed@tigr.org
Details of T. brucei sequencing at the Sanger Centre are available
at http://www.sanger.ac.uk/projects/T\_brucei/.
Location/Qualifiers
1. .541
/organism="Trypanosoma brucei"
/strain="TREU927"
/db_xref="taxon:5691"
/clone="249a12"

```

```

BASE COUNT 101 a 154 c 142 g 144 t
ORIGIN

```

```

Query Match 74.4%; Score 18.6; DB 17; Length 541;

```

Best Local Similarity 84.0%; Pred. No. 1e+03; Mismatches 0; Gaps 0;
Matches 21; Conservative 0; Indels 4;

QY 1 CGCGTCGGTGCAGGCGTCACAAAT 25
Db 330 CGCGGCGGTGCAGGATCTGAAAAAT 306

RESULT 4
BG330319/c
LOCUS
DEFINITION 602430347f1 NIH_MGC_18 Homo sapiens cDNA clone IMAGE:4548275 5',
mRNA sequence.
ACCESSION BG330319
VERSION BG330319.1 GI:13136757
KEYWORDS EST.
SOURCE human.

ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 786)
TITLE NIH-MGC http://mgc.nci.nih.gov/
JOURNAL National Institutes of Health, Mammalian Gene Collection (MGC)
COMMENT Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-femail.nih.gov

Tissue Procurement: DCTD/DTF/Gazdar
cDNA Library Preparation: Ling Hong/Rubin Laboratory
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: L1CM1237 row: m column: 12
High quality sequence stop: 544.
Location/Qualifiers
1..786

FEATURES

source
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4548275"
/clone_lib="NIH_MGC_18"
/lab_host="DH10B (phage-resistant)"
/tissue_type="large cell
/note="Organ: lung; Vector: pOTB7; Site1: XhoI; Site2:
ECORI; cDNA made by oligo-dt priming. Directionally cloned
into ECORI/XhoI sites using the following 5' adaptor:
GACACAG(G). Library constructed by Ling Hong in the
laboratory of Gerald M. Rubin (University of California,
Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and
Superscript II RT (Life Technologies). Note: this is a
NIH_MGC Library."
BASE COUNT 138 a 266 c 247 g 135 t
ORIGIN

Query Match 74.4%; Score 18.6; DB 12; Length 786;
Best Local Similarity 84.0%; Pred. No. 1e+03;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 CGCGTCGGTGCAGGCGTCACAAAT 25
Db 700 CGCGGCGGTGCAGGCGTGCACAT 676

RESULT 5
AQ235963/c
LOCUS
DEFINITION HS_2056_AL_D06_07 CIT Approved Human Genomic Sperm Library D Homo
sapiens genomic clone Plate=2056 Col=11 Row=G, DNA sequence.
ACCESSION AQ235963
VERSION AQ235963.1 GI:3664570
KEYWORDS GSS.
SOURCE human.
ORGANISM Homo sapiens

REFERENCE

AUTHORS
TITLE
JOURNAL
MEDLINE
COMMENT

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 516)
Mahairas, G., Wallace, J.C., Smith, K., Swartzell, S., Holzman, T.,
Keller, A., Shaker, R., Furlong, J., Young, J., Zhao, S., Adams, M.D. and
Hood, L.
Sequence-tagged connectors: A sequence approach to mapping and
scanning the human genome
Proc. Natl. Acad. Sci. U.S.A. 96 (17), 9739-9744 (1999)
99380589
Contact: Mahairas GG, Wallace JC, Hood L
High Throughput Sequencing Center
University of Washington
401 Queen Anne Avenue North, Seattle, WA 98109, USA
Tel: (206) 616-3618
Fax: (206) 616-3887
Email: jwallace@u.washington.edu

Sequence Tagged Connector
Plate: 2056 row: G column: 11
Class: BAC ends
High quality sequence stop: 516.
Location/Qualifiers
1..516

/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="Plate=2056 Col=11 Row=G"
/clone_lib="CIT Approved Human Genomic Sperm Library D"
/sex="male"
/note="Organ: sperm; Vector: pBelobAC11; BAC Clones in
E-Coli DH10B"
BASE COUNT 149 a 100 c 92 g 168 t 7 others
ORIGIN

Query Match 71.2%; Score 17.8; DB 17; Length 516;
Best Local Similarity 82.6%; Pred. No. 2.2e+03;
Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 CGCGTCGGTGCAGGCGTCACAA 23
Db 27 CGCGTCGGTGCAGGTCACAA 5

RESULT 6
CNS04KEP/c
LOCUS
DEFINITION

CNS04KEP
865 bp DNA linear GSS 21-MAY-2000
Tetraodon nigroviridis genome survey sequence T7 end of clone
116114 of library G from Tetraodon nigroviridis, genomic survey
sequence.
AL294720
GSS: genome survey sequence.
Tetraodon nigroviridis.
Tetraodon nigroviridis
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
Acanthomorpha; Acanthopterygii; Percormorpha; Tetraodontiformes;
Tetraodontidae; Tetraodon.
1 (bases 1 to 865)
Roest-Crollius, H., Jaillon, O., Dasilva, C., Bouneau, L., Fisher, C.,
Bernot, A., Fizames, C., Wincker, P., Brottier, P., Quetier, F.,
Saurin, W. and Weissenbach, J.
Human gene number estimate provided by genome wide analysis using
Tetraodon nigroviridis DNA sequence
Unpublished
2 (bases 1 to 865)
Roest-Crollius, H., Jaillon, O., Dasilva, C., Fizames, C., Fisher, C.,
Bouneau, L., Billault, A., Quetier, F., Saurin, W., Bernot, A. and
Weissenbach, J.
Characterization and repeat analysis of the compact genome of the
freshwater pufferfish Tetraodon nigroviridis
Unpublished
3 (bases 1 to 865)
Genoscope.

FEATURES

source
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="Plate=2056 Col=11 Row=G"
/clone_lib="CIT Approved Human Genomic Sperm Library D"
/sex="male"
/note="Organ: sperm; Vector: pBelobAC11; BAC Clones in
E-Coli DH10B"
BASE COUNT 149 a 100 c 92 g 168 t 7 others
ORIGIN

FEATURES

source
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="Plate=2056 Col=11 Row=G"
/clone_lib="CIT Approved Human Genomic Sperm Library D"
/sex="male"
/note="Organ: sperm; Vector: pBelobAC11; BAC Clones in
E-Coli DH10B"
BASE COUNT 149 a 100 c 92 g 168 t 7 others
ORIGIN

FEATURES

source
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="Plate=2056 Col=11 Row=G"
/clone_lib="CIT Approved Human Genomic Sperm Library D"
/sex="male"
/note="Organ: sperm; Vector: pBelobAC11; BAC Clones in
E-Coli DH10B"
BASE COUNT 149 a 100 c 92 g 168 t 7 others
ORIGIN

FEATURES

source
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="Plate=2056 Col=11 Row=G"
/clone_lib="CIT Approved Human Genomic Sperm Library D"
/sex="male"
/note="Organ: sperm; Vector: pBelobAC11; BAC Clones in
E-Coli DH10B"
BASE COUNT 149 a 100 c 92 g 168 t 7 others
ORIGIN

FEATURES

source
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="Plate=2056 Col=11 Row=G"
/clone_lib="CIT Approved Human Genomic Sperm Library D"
/sex="male"
/note="Organ: sperm; Vector: pBelobAC11; BAC Clones in
E-Coli DH10B"
BASE COUNT 149 a 100 c 92 g 168 t 7 others
ORIGIN

FEATURES

source
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="Plate=2056 Col=11 Row=G"
/clone_lib="CIT Approved Human Genomic Sperm Library D"
/sex="male"
/note="Organ: sperm; Vector: pBelobAC11; BAC Clones in
E-Coli DH10B"
BASE COUNT 149 a 100 c 92 g 168 t 7 others
ORIGIN

FEATURES

source
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="Plate=2056 Col=11 Row=G"
/clone_lib="CIT Approved Human Genomic Sperm Library D"
/sex="male"
/note="Organ: sperm; Vector: pBelobAC11; BAC Clones in
E-Coli DH10B"
BASE COUNT 149 a 100 c 92 g 168 t 7 others
ORIGIN

FEATURES

source
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="Plate=2056 Col=11 Row=G"
/clone_lib="CIT Approved Human Genomic Sperm Library D"
/sex="male"
/note="Organ: sperm; Vector: pBelobAC11; BAC Clones in
E-Coli DH10B"
BASE COUNT 149 a 100 c 92 g 168 t 7 others
ORIGIN

FEATURES

source
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="Plate=2056 Col=11 Row=G"
/clone_lib="CIT Approved Human Genomic Sperm Library D"
/sex="male"
/note="Organ: sperm; Vector: pBelobAC11; BAC Clones in
E-Coli DH10B"
BASE COUNT 149 a 100 c 92 g 168 t 7 others
ORIGIN

FEATURES

source
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="Plate=2056 Col=11 Row=G"
/clone_lib="CIT Approved Human Genomic Sperm Library D"
/sex="male"
/note="Organ: sperm; Vector: pBelobAC11; BAC Clones in
E-Coli DH10B"
BASE COUNT 149 a 100 c 92 g 168 t 7 others
ORIGIN

FEATURES

source
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="Plate=2056 Col=11 Row=G"
/clone_lib="CIT Approved Human Genomic Sperm Library D"
/sex="male"
/note="Organ: sperm; Vector: pBelobAC11; BAC Clones in
E-Coli DH10B"
BASE COUNT 149 a 100 c 92 g 168 t 7 others
ORIGIN

FEATURES

source
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="Plate=2056 Col=11 Row=G"
/clone_lib="CIT Approved Human Genomic Sperm Library D"
/sex="male"
/note="Organ: sperm; Vector: pBelobAC11; BAC Clones in
E-Coli DH10B"
BASE COUNT 149 a 100 c 92 g 168 t 7 others
ORIGIN

FEATURES

source
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="Plate=2056 Col=11 Row=G"
/clone_lib="CIT Approved Human Genomic Sperm Library D"
/sex="male"
/note="Organ: sperm; Vector: pBelobAC11; BAC Clones in
E-Coli DH10B"
BASE COUNT 149 a 100 c 92 g 168 t 7 others
ORIGIN

FEATURES

source
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="Plate=2056 Col=11 Row=G"
/clone_lib="CIT Approved Human Genomic Sperm Library D"
/sex="male"
/note="Organ: sperm; Vector: pBelobAC11; BAC Clones in
E-Coli DH10B"
BASE COUNT 149 a 100 c 92 g 168 t 7 others
ORIGIN

FEATURES

source
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="Plate=2056 Col=11 Row=G"
/clone_lib="CIT Approved Human Genomic Sperm Library D"
/sex="male"
/note="Organ: sperm; Vector: pBelobAC11; BAC Clones in
E-Coli DH10B"
BASE COUNT 149 a 100 c 92 g 168 t 7 others
ORIGIN

FEATURES

source
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="Plate=2056 Col=11 Row=G"
/clone_lib="CIT Approved Human Genomic Sperm Library D"
/sex="male"
/note="Organ: sperm; Vector: pBelobAC11; BAC Clones in
E-Coli DH10B"
BASE COUNT 149 a 100 c 92 g 168 t 7 others
ORIGIN

FEATURES

source
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="Plate=2056 Col=11 Row=G"
/clone_lib="CIT Approved Human Genomic Sperm Library D"
/sex="male"
/note="Organ: sperm; Vector: pBelobAC11; BAC Clones in
E-Coli DH10B"
BASE COUNT 149 a 100 c 92 g 168 t 7 others
ORIGIN

FEATURES

source
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="Plate=2056 Col=11 Row=G"
/clone_lib="CIT Approved Human Genomic Sperm Library D"
/sex="male"
/note="Organ: sperm; Vector: pBelobAC11; BAC Clones in
E-Coli DH10B"
BASE COUNT 149 a 100 c 92 g 168 t 7 others
ORIGIN

FEATURES

source
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="Plate=2056 Col=11 Row=G"
/clone_lib="CIT Approved Human Genomic Sperm Library D"
/sex="male"
/note="Organ: sperm; Vector: pBelobAC11; BAC Clones in
E-Coli DH10B"
BASE COUNT 149 a 100 c 92 g 168 t 7 others
ORIGIN

FEATURES

source
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="Plate=2056 Col=11 Row=G"
/clone_lib="CIT Approved Human Genomic Sperm Library D"
/sex="male"
/note="Organ: sperm; Vector: pBelobAC11; BAC Clones in
E-Coli DH10B"
BASE COUNT 149 a 100 c 92 g 168 t 7 others
ORIGIN

FEATURES

source
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="Plate=2056 Col=11 Row=G"
/clone_lib="CIT Approved Human Genomic Sperm Library D"
/sex="male"
/note="Organ: sperm; Vector: pBelobAC11; BAC Clones in
E-Coli DH10B"
BASE COUNT 149 a 100 c 92 g 168 t 7 others
ORIGIN

Query Match 71.2%; Score 17.8; DB 17; Length 1031;
Best Local Similarity 90.5%; Pred. No. 2.4e+03;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 CCGCTCGGTGCGAGGACGTGAC 21
||||||| ||||||| |||
Db 665 CCGCTCGGAGCAGGACGGGAC 645

RESULT 8
BE190622 199 bp mRNA linear EST 04-DEC-2001
LOCUS Gm-c1037-2736 5', mRNA sequence.
DEFINITION Gm-c1037-2736 5', mRNA sequence.
ACCESSION BE190622
VERSION BE190622.1 GI:8669515
KEYWORDS EST.
SOURCE soybean.
ORGANISM Glycine max
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae;
Glycine.

REFERENCE 1 (bases 1 to 199)
AUTHORS Shoemaker,R., Keim,P., Vodkin,L., Erpelding,J., Corvelli,V., Khanna
A., Bolla,B., Marra,M., Hillier,L., Kucaba,T., Martin,J., Beck,C.,
Wyllie,T., Underwood,K., Steptoe,M., Theising,B., Allen,M., Bowers
Y., Person,B., Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurr
R., Ritter,E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann
R., Witterston,R. and Wilson,R.
Public Soybean EST Project
Unpublished (1999)
Contact: Shoemaker R/Public Soybean EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
Putative full length read
vector to vector length is 200 This clone is available through:
ResGen, Invitrogen Corp. 2130 South Memorial Parkway Huntsville, AL
35801 For further information call: (800)-533-4363 or contact via
email: ccu@resgen.com
Insert length: 817 Std Error: 0.00.
Location/Qualifiers
1.199
/organism="Glycine max"
/db_xref="taxon:3847"
/clone="GENOME SYSTEMS CLONE ID: Gm-c1037-2736"
/clone_lib="Gm-c1037"
/tissue_type="fully expanded leaves of greenhouse grown
plants"
/dev_stage="2 week old"
/lab_host="DH10B"
/note="vector: pSPORT1; Site_1: NotI; Site_2: SalI; This
cDNA library was constructed from mRNA isolated from fully
expanded leaves of greenhouse grown plants that were 2
weeks old. The library was prepared using the Life
Technologies pSuperScript cDNA library construction kit.
Complementary DNA was synthesized from mRNA using a
poly(dT) sequence with a NotI restriction site. SalI
linkers adapters were ligated to the blunt-ended cDNA
fragments followed by NotI digestion. The cDNA fragments
were directionally cloned into the NotI-SalI restriction
site of the pSPORT1 vector. The ligated cDNA fragments
were transformed into E.coli Electro-Max DH10B host
cells. This library was constructed in the laboratory of
Dr. Lila Vodkin by Anu Khanna at the University of
Illinois at Urbana- Champaign. email: l-vodkin@uiuc.edu"

BASE COUNT 39 a 35 c 49 g 76 t
ORIGIN

Direct Submission
Submitted (12-APR-2000)
This sequence is a single read and was generated as part of a large
scale clone-end sequencing project of the Tetraodon nigroviridis
genome. For more information, please take a look at
<http://www.genoscope.cns.fr/Tetraodon>.
Location/Qualifiers
1.865
/organism="Tetraodon nigroviridis"
/db_xref="taxon:99883"
/clone="116114"
/clone_lib="g"
/note="Genoscope sequence ID : COBG116BE07LP1-end : T7"

BASE COUNT 188 a 234 c 211 g 227 t 5 others
ORIGIN

Query Match 71.2%; Score 17.8; DB 17; Length 865;
Best Local Similarity 90.5%; Pred. No. 2.4e+03;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 CCGCTCGGTGCGAGGACGTGAC 21
||||||| ||||||| |||
Db 235 CCGCTCGGAGCAGGACGGGAC 215

RESULT 7
CNS04F4W/C 1031 bp DNA linear GSS 21-MAY-2000
LOCUS Tetraodon nigroviridis genome survey sequence PUC-Ori end of clone
DEFINITION 105J24 of library G from Tetraodon nigroviridis, genomic survey
sequence.
ACCESSION AL287897
VERSION AL287897.1 GI:8026407
KEYWORDS GSS; genome survey sequence.
SOURCE Tetraodon nigroviridis.
ORGANISM Tetraodon nigroviridis
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
Tetraodontidae; Tetraodon.
1 (bases 1 to 1031)
Roest-Crollius,H., Jaillon,O., Dasilva,C., Bouneau,L., Fisher,C.,
Bernot,A., Fizames,C., Wincker,P., Brottier,P., Quetier,F.,
Saurin,W. and Weissenbach,J.
Human gene number estimate provided by genome wide analysis using
Tetraodon nigroviridis DNA sequence
Unpublished
2 (bases 1 to 1031)
Roest-Crollius,H., Jaillon,O., Dasilva,C., Fizames,C., Fisher,C.,
Bouneau,L., Billault,A., Quetier,F., Saurin,W., Bernot,A. and
Weissenbach,J.
Characterization and repeat analysis of the compact genome of the
freshwater pufferfish Tetraodon nigroviridis
Unpublished
3 (bases 1 to 1031)
Genoscope.
Direct Submission
Submitted (12-APR-2000)
This sequence is a single read and was generated as part of a large
scale clone-end sequencing project of the Tetraodon nigroviridis
genome. For more information, please take a look at
<http://www.genoscope.cns.fr/Tetraodon>.
Location/Qualifiers
1.1031
/organism="Tetraodon nigroviridis"
/db_xref="taxon:99883"
/clone="105J24"
/clone_lib="g"
/note="Genoscope sequence ID : COBG105DE12SP1-end :
PUC-Ori"

BASE COUNT 219 a 282 c 311 g 197 t 22 others
ORIGIN

Query Match 70.4%; Score 17.6; DB 10; Length 199;
Best Local Similarity 83.3%; Pred. No. 2.2e+03;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 CGCGTGGTGCAGGACGTGACAAA 24
||||| ||| | ||| |||||
Db 4 CGCGTGGTGCAGGACGTGACAAA 27

RESULT 9

AQ647334/c

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

AQ647334
RPC193-ECORI-2B12.TV RPC193-ECORI Trypanosoma brucei genomic clone
RPC193-ECORI-2B12, DNA sequence.
AQ647334
GI:5140520
Trypanosoma brucei.
Trypanosoma brucei
Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae;
Trypanosoma.
1 (bases 1 to 412)
El-Sayed,N., Zhao,S., Zhao,H., Gill,S., Suh,E., Malek,J., Fujii,C.,
Gerrard,C., Leech,V., de Jong,P., Ullu,E., Melville,S., Donelson,J.,
Fraser,C. and Adams,M.
Use of BAC end sequences from Trypanosoma brucei GUTat 10.1 RPCI-93
Library for gene discovery and sequence-ready map construction
Unpublished (1999)
Other_GSSs: RPC193-ECORI-2B12.TP
Contact: Najib M. El-Sayed
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: nelsayed@tigr.org
Clones and high density filters may be purchased from BACPAC
Resources (http://bacpac.med.buffalo.edu). BAC end sequences search
page: http://www.tigr.org/tdb/mbd/tbdb/.
Seq primer: T7
Class: BAC ends.

FEATURES
source

1. .412
/organism="Trypanosoma brucei"
/strain="TREU927/4 GUTat 10.1"
/db_xref="taxon:5691"
/clone="RPC193-ECORI-2B12"
/clone_lib="RPC193-ECORI"
/notes="Vector: pBACe3.6; Site_1: Eco RI; Site_2: Eco RI;
Constructed for The Institute for Genomic Research by
Bohui Zhao in Pieter de Jong's laboratory (Roswell Park
Cancer Institute, Buffalo, NY). Briefly, Trypanosoma
brucei TREU927/4 GUTat 10.1 agarose embedded DNA was
partially digested with a combination of Eco RI and Eco RI
methylase (RPC193-ECORI segment) or Dpn II (RPC193-DpnII
segment). High molecular weight fragments were ligated in
pBACe3.6 vector digested with Eco RI or Bam HI,
respectively. The average insert size is 141 Kb. Total
coverage (both segments): > 90 x the haploid
non-minichromosomal genome."
105 a 115 c 95 g 97 t

Query Match 70.4%; Score 17.6; DB 17; Length 412;
Best Local Similarity 83.3%; Pred. No. 2.5e+03;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 CGCGTGGTGCAGGACGTGACAAA 24
||||| ||| | ||| |||||
Db 231 CGCGCCAGTGGAGGACGTGACAAA 208

RESULT 10

Query Match 70.4%; Score 17.6; DB 17; Length 414;
Best Local Similarity 83.3%; Pred. No. 2.5e+03;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 CGCGTGGTGCAGGACGTGACAAA 24
||||| ||| | ||| |||||
Db 184 CGCGCCAGTGGAGGACGTGACAAA 161

Query Match 70.4%; Score 17.6; DB 17; Length 414;
Best Local Similarity 83.3%; Pred. No. 2.5e+03;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 CGCGTGGTGCAGGACGTGACAAA 24
||||| ||| | ||| |||||
Db 184 CGCGCCAGTGGAGGACGTGACAAA 161

Query Match 70.4%; Score 17.6; DB 17; Length 414;
Best Local Similarity 83.3%; Pred. No. 2.5e+03;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 CGCGTGGTGCAGGACGTGACAAA 24
||||| ||| | ||| |||||
Db 184 CGCGCCAGTGGAGGACGTGACAAA 161

Query Match 70.4%; Score 17.6; DB 17; Length 414;
Best Local Similarity 83.3%; Pred. No. 2.5e+03;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 CGCGTGGTGCAGGACGTGACAAA 24
||||| ||| | ||| |||||
Db 184 CGCGCCAGTGGAGGACGTGACAAA 161

Query Match 70.4%; Score 17.6; DB 17; Length 414;
Best Local Similarity 83.3%; Pred. No. 2.5e+03;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 CGCGTGGTGCAGGACGTGACAAA 24
||||| ||| | ||| |||||
Db 184 CGCGCCAGTGGAGGACGTGACAAA 161

Query Match 70.4%; Score 17.6; DB 17; Length 414;
Best Local Similarity 83.3%; Pred. No. 2.5e+03;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 CGCGTGGTGCAGGACGTGACAAA 24
||||| ||| | ||| |||||
Db 184 CGCGCCAGTGGAGGACGTGACAAA 161

Query Match 70.4%; Score 17.6; DB 17; Length 414;
Best Local Similarity 83.3%; Pred. No. 2.5e+03;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 CGCGTGGTGCAGGACGTGACAAA 24
||||| ||| | ||| |||||
Db 184 CGCGCCAGTGGAGGACGTGACAAA 161

Query Match 70.4%; Score 17.6; DB 17; Length 414;
Best Local Similarity 83.3%; Pred. No. 2.5e+03;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 CGCGTGGTGCAGGACGTGACAAA 24
||||| ||| | ||| |||||
Db 184 CGCGCCAGTGGAGGACGTGACAAA 161

Query Match 70.4%; Score 17.6; DB 17; Length 414;
Best Local Similarity 83.3%; Pred. No. 2.5e+03;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 CGCGTGGTGCAGGACGTGACAAA 24
||||| ||| | ||| |||||
Db 184 CGCGCCAGTGGAGGACGTGACAAA 161

Query Match 70.4%; Score 17.6; DB 17; Length 414;
Best Local Similarity 83.3%; Pred. No. 2.5e+03;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 CGCGTGGTGCAGGACGTGACAAA 24
||||| ||| | ||| |||||
Db 184 CGCGCCAGTGGAGGACGTGACAAA 161

Query Match 70.4%; Score 17.6; DB 17; Length 414;
Best Local Similarity 83.3%; Pred. No. 2.5e+03;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 CGCGTGGTGCAGGACGTGACAAA 24
||||| ||| | ||| |||||
Db 184 CGCGCCAGTGGAGGACGTGACAAA 161

Query Match 70.4%; Score 17.6; DB 17; Length 414;
Best Local Similarity 83.3%; Pred. No. 2.5e+03;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 CGCGTGGTGCAGGACGTGACAAA 24
||||| ||| | ||| |||||
Db 184 CGCGCCAGTGGAGGACGTGACAAA 161

Query Match 70.4%; Score 17.6; DB 17; Length 414;
Best Local Similarity 83.3%; Pred. No. 2.5e+03;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 CGCGTGGTGCAGGACGTGACAAA 24
||||| ||| | ||| |||||
Db 184 CGCGCCAGTGGAGGACGTGACAAA 161

Query Match 70.4%; Score 17.6; DB 17; Length 414;
Best Local Similarity 83.3%; Pred. No. 2.5e+03;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 CGCGTGGTGCAGGACGTGACAAA 24
||||| ||| | ||| |||||
Db 184 CGCGCCAGTGGAGGACGTGACAAA 161

Query Match 70.4%; Score 17.6; DB 17; Length 414;
Best Local Similarity 83.3%; Pred. No. 2.5e+03;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 CGCGTGGTGCAGGACGTGACAAA 24
||||| ||| | ||| |||||
Db 184 CGCGCCAGTGGAGGACGTGACAAA 161

Query Match 70.4%; Score 17.6; DB 17; Length 414;
Best Local Similarity 83.3%; Pred. No. 2.5e+03;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 CGCGTGGTGCAGGACGTGACAAA 24
||||| ||| | ||| |||||
Db 184 CGCGCCAGTGGAGGACGTGACAAA 161

Query Match 70.4%; Score 17.6; DB 17; Length 414;
Best Local Similarity 83.3%; Pred. No. 2.5e+03;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 CGCGTGGTGCAGGACGTGACAAA 24
||||| ||| | ||| |||||
Db 184 CGCGCCAGTGGAGGACGTGACAAA 161

Query Match 70.4%; Score 17.6; DB 17; Length 414;
Best Local Similarity 83.3%; Pred. No. 2.5e+03;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 CGCGTGGTGCAGGACGTGACAAA 24
||||| ||| | ||| |||||
Db 184 CGCGCCAGTGGAGGACGTGACAAA 161

Query Match 70.4%; Score 17.6; DB 17; Length 414;
Best Local Similarity 83.3%; Pred. No. 2.5e+03;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 CGCGTGGTGCAGGACGTGACAAA 24
||||| ||| | ||| |||||
Db 184 CGCGCCAGTGGAGGACGTGACAAA 161

Query Match 70.4%; Score 17.6; DB 17; Length 414;
Best Local Similarity 83.3%; Pred. No. 2.5e+03;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 CGCGTGGTGCAGGACGTGACAAA 24
||||| ||| | ||| |||||
Db 184 CGCGCCAGTGGAGGACGTGACAAA 161

Query Match 70.4%; Score 17.6; DB 17; Length 414;
Best Local Similarity 83.3%; Pred. No. 2.5e+03;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 CGCGTGGTGCAGGACGTGACAAA 24
||||| ||| | ||| |||||
Db 184 CGCGCCAGTGGAGGACGTGACAAA 161

Query Match 70.4%; Score 17.6; DB 17; Length 414;
Best Local Similarity 83.3%; Pred. No. 2.5e+03;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 CGCGTGGTGCAGGACGTGACAAA 24
||||| ||| | ||| |||||
Db 184 CGCGCCAGTGGAGGACGTGACAAA 161

Query Match 70.4%; Score 17.6; DB 17; Length 414;
Best Local Similarity 83.3%; Pred. No. 2.5e+03;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 CGCGTGGTGCAGGACGTGACAAA 24
||||| ||| | ||| |||||
Db 184 CGCGCCAGTGGAGGACGTGACAAA 161

AQ844401
LOCUS an36h09 JM101 filtered library Zea mays genomic, DNA sequence.
ACCESSION AQ844401
VERSION AQ844401.1 GI:6202889
KEYWORDS GSS.
SOURCE Zea mays.
ORGANISM Zea mays.
REFERENCE
AUTHORS Rabinowicz,P.D., Schutz,K., Dedhia,N., Yordan,C., Parnell,L.D., Stein,L., McCombie,W.R. and Martienssen,R.A.
TITLE Differential methylation of genes and retrotransposons allows shotgun sequencing of the maize genome
JOURNAL Nat. Genet. 23, 305-308 (1999)
COMMENT Contact: Martienssen RA
Cold Spring Harbor Laboratory
1 Bungtown Rd., Cold Spring Harbor, NY 11724, USA
Tel: 516 367 8322
Fax: 516 367 8369
Email: martiensscshl.org
Seq primer: forward
Class: shotgun.

FEATURES
source
1. 464
/organism="Zea mays"
/cultivar="B73"
/db_xref="taxon:4577"
/clone_lib="JM101 filtered library"
/note="Organ: immature ears; Vector: M13; Site_1: Xba I;
DNA prepared from purified nuclei was digested with the
methylation insensitive enzyme Spe I, size fractionated to
enrich for the 0.5 to 4 kbp fraction, ligated into Xba I
digested M13 vector and electroporated into E.coli JM101."
BASE COUNT 77 a 146 c 156 g 85 t
ORIGIN
Query Match 70.4%; Score 17.6; DB 17; Length 464;
Best Local Similarity 83.3%; Pred. No. 2.6e+03;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
QY 1 CGCGTCGTCGACGACGTGCATA 24
||||| ||||||||| |||
Db 149 CGCGTCGTCGACGACGTGCATA 172
RESULT 14
AA957358 515 bp mRNA linear EST 04-JUL-1999
LOCUS UI-R-EI-fy-b-07-0-UI.s1 UI-R-EI Rattus norvegicus cDNA clone
DEFINITION UI-R-EI-fy-b-07-0-UI 3', mRNA sequence.
ACCESSION AA957358
VERSION AA957358.1 GI:4277248
KEYWORDS EST.
SOURCE Norway rat.
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
1 (bases 1 to 515)
REFERENCE
AUTHORS Ronaldo,M.F., Lennon,G. and Soares,M.B.
TITLE Normalization and subtraction: two approaches to facilitate gene
discovery
JOURNAL Genome Res. 6 (9), 791-806 (1996)
MEDLINE 97044477
COMMENT On May 7, 1998 this sequence version replaced gi:3121053.
Contact: Soares, MB
Program for Rat Gene Discovery and Mapping
University of Iowa
451 Eckstein Medical Research Building Iowa City, IA 52242, USA

AQ844401
LOCUS an36h09 JM101 filtered library Zea mays genomic, DNA sequence.
ACCESSION AQ844401
VERSION AQ844401.1 GI:6202889
KEYWORDS GSS.
SOURCE Zea mays.
ORGANISM Zea mays.
REFERENCE
AUTHORS Rabinowicz,P.D., Schutz,K., Dedhia,N., Yordan,C., Parnell,L.D., Stein,L., McCombie,W.R. and Martienssen,R.A.
TITLE Differential methylation of genes and retrotransposons allows shotgun sequencing of the maize genome
JOURNAL Nat. Genet. 23, 305-308 (1999)
COMMENT Contact: Martienssen RA
Cold Spring Harbor Laboratory
1 Bungtown Rd., Cold Spring Harbor, NY 11724, USA
Tel: 516 367 8322
Fax: 516 367 8369
Email: martiensscshl.org
Seq primer: forward
Class: shotgun.

FEATURES
source
1. 464
/organism="Zea mays"
/cultivar="B73"
/db_xref="taxon:4577"
/clone_lib="JM101 filtered library"
/note="Organ: immature ears; Vector: M13; Site_1: Xba I;
DNA prepared from purified nuclei was digested with the
methylation insensitive enzyme Spe I, size fractionated to
enrich for the 0.5 to 4 kbp fraction, ligated into Xba I
digested M13 vector and electroporated into E.coli JM101."
BASE COUNT 77 a 146 c 156 g 85 t
ORIGIN
Query Match 70.4%; Score 17.6; DB 17; Length 464;
Best Local Similarity 83.3%; Pred. No. 2.6e+03;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
QY 1 CGCGTCGTCGACGACGTGCATA 24
||||| ||||||||| |||
Db 149 CGCGTCGTCGACGACGTGCATA 172
RESULT 14
AA957358 515 bp mRNA linear EST 04-JUL-1999
LOCUS UI-R-EI-fy-b-07-0-UI.s1 UI-R-EI Rattus norvegicus cDNA clone
DEFINITION UI-R-EI-fy-b-07-0-UI 3', mRNA sequence.
ACCESSION AA957358
VERSION AA957358.1 GI:4277248
KEYWORDS EST.
SOURCE Norway rat.
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
1 (bases 1 to 515)
REFERENCE
AUTHORS Ronaldo,M.F., Lennon,G. and Soares,M.B.
TITLE Normalization and subtraction: two approaches to facilitate gene
discovery
JOURNAL Genome Res. 6 (9), 791-806 (1996)
MEDLINE 97044477
COMMENT On May 7, 1998 this sequence version replaced gi:3121053.
Contact: Soares, MB
Program for Rat Gene Discovery and Mapping
University of Iowa
451 Eckstein Medical Research Building Iowa City, IA 52242, USA

Tel: 319 335 8250
Fax: 319 335 9565
Email: msoares@blue.weeg.uiowa.edu
Oligo-dT track not found, Not 1 site shown in beginning of sequence.
is likely internal to the message. cDNA library preparation: M.
Fatima Bonaldo, Ph.D. Clone distribution: clones will be available
through Research Genetics This clone is also available through the
I.M.A.G.E. Consortium at LLNL (info@image.llnl.gov). IMAGE
ID=1780190 The following repetitive elements were found in this
cDNA sequence: 57-96, >GC-richLow_complexity
Seq primer: M13 Forward
POLYA-No. Location/Qualifiers
1. 515
/organism="Rattus norvegicus"
/strain="Sprague-Dawley"
/db_xref="taxon:10116"
/clone="UI-R-EI-fy-b-07-0-UI"
/clone_lib="UI-R-EI"
/dev_stage="adult"
/lab_host="DH10B (Life Technologies)"
/note="Vector: pT73D-pac (Pharmacia) with a modified
polylinker: Site_1: Not I; Site_2: Eco RI; The UI-R-EI
library is a subtracted library derived from the UI-R-EI
library. The UI-R-EI library consisted of a mixture of
individually tagged normalized libraries constructed from
8, 12 and 18-day embryo. The tag is a string of 3-5
nucleotides present between the Not I site and the
oligo-dT track which allows identification of the library
of origin of a clone within the mixture. The subtracted
library (UI-R-EI) was constructed as follows: PCR
amplified cDNA inserts from a pool of UI-R-EI clones from
which 3' ESTs had been derived was used as a driver in a
hybridization with the UI-R-EI library in the form of
single-stranded circles. The remaining single-stranded
circles (subtracted library) was purified by
hydroxyapatite column chromatography, converted to
double-stranded circles and electroporated into DH10B
bacteria (Life Technologies) to generate the UI-R-EI
library. This procedure has been previously described
(Bonaldo, Lennon and Soares, Genome Research 6: 791-806,
1996)."
BASE COUNT 159 a 116 c 149 g 91 t
ORIGIN
Query Match 70.4%; Score 17.6; DB 9; Length 515;
Best Local Similarity 83.3%; Pred. No. 2.6e+03;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
QY 1 CGCGTCGTCGACGACGTGCACAA 24
||||| ||||||||| |||
Db 467 CACGTCGTCGACGACGTGCACAA 490
RESULT 15
AW664262/c 564 bp mRNA linear EST 06-APR-2000
LOCUS h108e10.x1 NCI_CGAP_GUI Homo sapiens cDNA clone IMAGE:2971722 3'
DEFINITION similar to FR:O35540 O35540 HEPATOMA-DERIVED GROWTH FACTOR, RELATED
PROTEIN 2. ; mRNA sequence.
ACCESSION AW664262
VERSION AW664262.1 GI:7456803
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 564)
REFERENCE
AUTHORS NCI_CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.

Mon Jan 6 15:20:17 2003

Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Chris Moskaluk, M.D., Ph.D., Michael R. Emmert-Buck, M.D., Ph.D. cDNA Library Preparation: Life Technologies, Inc. cDNA Library Arrayed by: Christa Prange, The I.M.A.G.E. Consortium DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov/image/html/iresources.shtml

Possible reversed clone: polyt not found

Seq primer: -40UP from Gibco

High quality sequence stop: 398.

FEATURES

source

Location/Qualifiers

1..564

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="IMAGE:2971722"

/clone_lib="NCI-CGAP_GUI"

/tissue_type="2 pooled high-grade transitional cell

tumors"

/lab_host="DH10B"

/note="Organ: genitourinary tract; Vector: pCMV-SPORT6;

Site_1: SalI; Site_2: NotI; Cloned unidirectionally.

Primer: Oligo dT. Library constructed by Life

Technologies."

BASE COUNT

85 a 202 c 134 g 142 t 1 others

ORIGIN

Query Match

Best Local Similarity 70.4%; Score 17.6; DB 10; Length 564;

Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 CGCGTCGGTGCAGGACGTGACAAA 24

||| ||||| ||||| |||||

Db 562 CGAGCGGTGCAGAAAGTGACAAA 539

Search completed: January 4, 2003, 01:04:07

Job time : 338.413 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 3, 2003, 22:53:16 ; Search time 8.43849 Seconds

(without alignments)
908.566 Million cell updates/sec

Title: US-09-787-562-1

Perfect score: 25

Sequence: 1 cgcgcggtgcaggacgtgacaaat 25

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 441362 seqs, 153338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Issued_Patents_NA:*
1: /cgn2_6/ptodata/2/ina/5A.COMB.seq.*
2: /cgn2_6/ptodata/2/ina/5B.COMB.seq.*
3: /cgn2_6/ptodata/2/ina/6A.COMB.seq.*
4: /cgn2_6/ptodata/2/ina/6B.COMB.seq.*
5: /cgn2_6/ptodata/2/ina/PTUS.COMB.seq.*
6: /cgn2_6/ptodata/2/ina/backfiles1.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	18.6	74.4	3670	1	US-08-386-727-3
2	18.6	74.4	3670	2	US-08-600-452A-3
3	18	72.0	24	2	US-08-693-174-2
4	18	72.0	24	4	US-09-253-738-2
5	18	72.0	1110	2	US-08-693-174-4
6	18	72.0	1110	4	US-09-253-738-4
7	18	72.0	5382	4	US-09-479-122-21
8	18	72.0	7617	3	US-08-646-538-34
9	18	72.0	7617	4	US-09-503-222-34
10	18	72.0	8387	2	US-08-532-814-1
11	18	72.0	8388	4	US-09-225-509-1
12	18	72.0	9737	4	US-09-479-122-22
13	18	72.0	9737	4	US-09-479-122-23
14	18	72.0	9737	4	US-09-479-122-28
15	18	72.0	9871	4	US-09-479-122-24
16	18	72.0	10060	4	US-09-479-122-25
17	17.6	70.4	4256	1	US-08-505-509-31
18	17.6	70.4	4256	2	US-08-491-690A-31
19	16.2	64.8	6671	1	US-08-280-443-1
20	16.2	64.8	6671	1	US-08-457-459-1
21	16.2	64.8	6671	1	US-08-555-678-1
22	16.2	64.8	6671	5	PCT-US95-02275-1
23	16	64.0	1464	4	US-09-351-224E-4
24	16	64.0	2353	5	PCT-US92-06840-1
25	16	64.0	3735	4	US-08-975-762-43
26	16	64.0	3735	4	US-09-295-028-43
27	16	64.0	3735	4	US-09-106-582-43

28 16 64.0 12847 1 US-08-550-715-1 Sequence 1, Appli
29 16 64.0 4411529 4 US-09-103-840A-1 Sequence 1, Appli
30 15.8 63.2 372 4 US-09-124-671-24 Sequence 24, Appli
31 15.8 63.2 1815 3 US-09-041-545-1 Sequence 1, Appli
32 15.8 63.2 1815 3 US-09-327-925-1 Sequence 1, Appli
33 15.8 63.2 18627 4 US-08-961-527-113 Sequence 113, App
34 15.8 63.2 43950 4 US-09-735-934A-3 Sequence 3, Appli
35 15.8 63.2 4403765 4 US-09-103-840A-2 Sequence 2, Appli
36 15.6 62.4 200 1 US-08-308-892A-13 Sequence 13, Appli
37 15.6 62.4 200 1 US-08-308-892A-14 Sequence 14, Appli
38 15.6 62.4 200 1 US-08-308-892A-15 Sequence 15, Appli
39 15.6 62.4 274 4 US-08-990-823-72 Sequence 72, Appli
40 15.6 62.4 507 1 US-08-246-427A-1 Sequence 1, Appli
41 15.6 62.4 507 2 US-08-766-620-1 Sequence 1, Appli
42 15.6 62.4 507 5 PCT-US95-06094-1 Sequence 1, Appli
43 15.6 62.4 525 1 US-08-009-973-2 Sequence 2, Appli
44 15.6 62.4 540 4 US-08-149-101A-3 Sequence 3, Appli
45 15.6 62.4 540 5 PCT-US94-12873-3 Sequence 3, Appli

ALIGNMENTS

RESULT 1
US-08-386-727-3
; Sequence 3, Application US/08386727
; Patent No. 5792647
; GENERAL INFORMATION:
; APPLICANT: ROSEMAN, SAUL
; APPLICANT: BASSLER, BONNIE
; APPLICANT: KEYHANT, NEMAT O.
; APPLICANT: CHITLARU, EDITH
; APPLICANT: ROWE, CHRIS
; APPLICANT: YU, CHARLES
; TITLE OF INVENTION: BACTERIAL CATABOLISM OF CHITIN
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: CUSHMAN, DARBY & CUSHMAN
; STREET: 1100 NEW YORK AVENUE, N.W.
; CITY: WASHINGTON
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20005
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/386,727
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: HOBBS, ANN S.
; REGISTRATION NUMBER: 36,830
; REFERENCE/DOCKET NUMBER: 4130/206916
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-861-3000
; TELEFAX: 202-822-0944
; TELEX: 6714627 CUSH
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3670 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
US-08-386-727-3

Query Match 74.4%; Score 18.6; DB 1; Length 3670;
Best Local Similarity 84.0%; Pred. No. 7.9;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Mon Jan 6 15:20:17 2003

us-09-787-562-1.rni

Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 GTGCAGGACGTGACAAAT 25
 Db 637 GTGCAGGACGTGACAAAT 654

RESULT 6
 US-09-253-738-4
 ; Sequence 4, Application US/09253738
 ; Patent No. 6265390
 ; GENERAL INFORMATION:
 ; APPLICANT: Ratcliffe, Peter John
 ; APPLICANT: Firth, John David
 ; APPLICANT: Harris, Adrian Llewellyn
 ; APPLICANT: Pugh, Christopher William
 ; APPLICANT: Stratford, Ian James
 ; TITLE OF INVENTION: Targeting Gene Therapy
 ; FILE REFERENCE: 08/693174
 ; CURRENT APPLICATION NUMBER: US/09/253,738
 ; CURRENT FILING DATE: 1999-02-22
 ; NUMBER OF SEQ ID NOS: 5
 ; SOFTWARE: PatentIn Ver. 2.0
 ; SEQ ID NO 4
 ; LENGTH: 1110
 ; TYPE: DNA
 ; ORGANISM: Murinae gen. sp.
 US-09-253-738-4

Query Match 72.0%; Score 18; DB 4; Length 1110;
 Best Local Similarity 100.0%; Pred. No. 13;
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 GTGCAGGACGTGACAAAT 25
 Db 637 GTGCAGGACGTGACAAAT 654

RESULT 7
 US-09-479-122-21
 ; Sequence 21, Application US/09479122
 ; Patent No. 6410266
 ; GENERAL INFORMATION:
 ; APPLICANT: HARRINGTON, JOHN J.
 ; APPLICANT: SHERF, BRUCE
 ; APPLICANT: RUNDLETT, STEPHEN
 ; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR NON-TARGETED ACTIVATION OF
 ; FILE REFERENCE: 0221-0003C
 ; CURRENT APPLICATION NUMBER: US/09/479,122
 ; CURRENT FILING DATE: 2000-01-07
 ; PRIOR APPLICATION NUMBER: 09/276,820
 ; PRIOR FILING DATE: 1999-03-26
 ; PRIOR APPLICATION NUMBER: 09/159,643
 ; PRIOR FILING DATE: 1998-09-24
 ; PRIOR APPLICATION NUMBER: 08/941,223
 ; PRIOR FILING DATE: 1997-09-26
 ; PRIOR APPLICATION NUMBER: 09/263,814
 ; PRIOR FILING DATE: 1999-03-08
 ; PRIOR APPLICATION NUMBER: 09/253,022
 ; PRIOR FILING DATE: 1999-02-19
 ; NUMBER OF SEQ ID NOS: 33
 ; SOFTWARE: PatentIn Ver. 2.1
 ; SEQ ID NO 21
 ; LENGTH: 5382
 ; TYPE: DNA
 ; ORGANISM: Homo sapiens
 ; FEATURE:
 ; NAME/KEY: modified_base
 ; LOCATION: (890)
 ; OTHER INFORMATION: a, c, t, g, other or unknown
 ; NAME/KEY: modified_base
 ; LOCATION: (1042)

; OTHER INFORMATION: a, c, t, g, other or unknown
 US-09-479-122-21

Query Match 72.0%; Score 18; DB 4; Length 5382;
 Best Local Similarity 100.0%; Pred. No. 15;
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 GTGCAGGACGTGACAAAT 25
 Db 899 GTGCAGGACGTGACAAAT 916

RESULT 8
 US-08-646-538-34/C
 ; Sequence 34, Application US/08646538
 ; Patent No. 6027881
 ; GENERAL INFORMATION:
 ; APPLICANT: Pavlakis, George N.
 ; APPLICANT: Gaitanaris, George A.
 ; APPLICANT: Stauber, Roland H.
 ; APPLICANT: Vournakis, John N.
 ; TITLE OF INVENTION: Mutant Aequorea victoria Fluorescent
 ; NUMBER OF SEQUENCES: 37
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Townsend and Townsend and Crew LLP
 ; STREET: Two Embarcadero Center, 8th Floor
 ; CITY: San Francisco
 ; STATE: California
 ; COUNTRY: USA
 ; ZIP: 94111-3834
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC Compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: PatentIn Release #1.0, Version #1.30
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/646,538
 ; FILING DATE: No. 6027881 yet assigned
 ; CLASSIFICATION: 435
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Weber, Kenneth A.
 ; REGISTRATION NUMBER: 31,677
 ; REFERENCE/DOCKET NUMBER: 015280-249000
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (415) 576-0200
 ; TELEFAX: (415) 576-0300
 ; INFORMATION FOR SEQ ID NO: 34:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 7617 base pairs
 ; TYPE: nucleic acid
 ; STRANDEDNESS: single
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: DNA
 ; FEATURE:
 ; NAME/KEY: -
 ; LOCATION: 1..7617
 ; OTHER INFORMATION: /note= "pGen-PGK9fo25RO"
 US-08-646-538-34

Query Match 72.0%; Score 18; DB 3; Length 7617;
 Best Local Similarity 100.0%; Pred. No. 16;
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 GTGCAGGACGTGACAAAT 25
 Db 4186 GTGCAGGACGTGACAAAT 4169

RESULT 9
 US-09-503-222-34/c
 ; Sequence 34, Application US/09503222
 ; Patent No. 6265548

```

; GENERAL INFORMATION:
; APPLICANT: Pavlakis, George N.
; APPLICANT: Gaitanaris, George A.
; APPLICANT: Stauber, Roland H.
; APPLICANT: Vournakis, John N.
; TITLE OF INVENTION: Mutant Aequorea victoria Fluorescent
; TITLE OF INVENTION: Proteins Having Increased Cellular Fluorescence
; NUMBER OF SEQUENCES: 37
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Townsend and Townsend and Crew LLP
; STREET: Two Embarcadero Center, 8th Floor
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94111-3834
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/503,222
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/646,538
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Weber, Kenneth A.
; REGISTRATION NUMBER: 31,677
; REFERENCE/DOCKET NUMBER: 015280-249000
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 576-0200
; TELEFAX: (415) 576-0300
; INFORMATION FOR SEQ ID NO: 34:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 7617 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA
; FEATURE:
; NAME/KEY: -
; LOCATION: 1..7617
; OTHER INFORMATION: /note= "pGen-pCKgfo25RO"
; US-09-503-222-34

Query Match 72.0%; Score 18; DB 4; Length 7617;
Best Local Similarity 100.0%; Pred. No. 16;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 GTGCAGGACGTGACAAAT 25
Db 4186 GTGCAGGACGTGACAAAT 4169
|||||
; GENERAL INFORMATION:
; APPLICANT: MOULLIER, PHILIPPE
; APPLICANT: DANOS, OLIVIER
; APPLICANT: HEARD, JEAN-MICHEL
; APPLICANT: FERRY, NICOLAS
; TITLE OF INVENTION: BIOCOMPATIBLE IMPLANT FOR THE EXPRESSION
; TITLE OF INVENTION: AND IN VIVO SECRETION OF A THERAPEUTIC SUBSTANCE
; NUMBER OF SEQUENCES: 1
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, MAIER & NEUSTADT,
; ADDRESS: P.C.
; STREET: 1755 S. JEFFERSON DAVIS HIGHWAY, SUITE 400
; CITY: ARLINGTON

```

```

; STATE: VA
; COUNTRY: USA
; ZIP: 22202
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/532,814
; FILING DATE: 19-JAN-1996
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/FR94/00456
; FILING DATE: 21-APR-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: FR 93/04700
; FILING DATE: 21-APR-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: FR 93/09185
; FILING DATE: 26-JUL-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: OBLON, NORMAN F.
; REGISTRATION NUMBER: 24,618
; REFERENCE/DOCKET NUMBER: 660-105-0 PCT
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 703-413-3000
; TELEFAX: 703-413-2220
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 8387 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; US-08-532-814-1

Query Match 72.0%; Score 18; DB 2; Length 8387;
Best Local Similarity 100.0%; Pred. No. 16;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 GTGCAGGACGTGACAAAT 25
Db 2238 GTGCAGGACGTGACAAAT 2255
|||||
; GENERAL INFORMATION:
; APPLICANT: MOULLIER, Philippe
; APPLICANT: DANOS, Olivier
; APPLICANT: HEARD, Jean-Michel
; APPLICANT: FERRY, Nicholas
; TITLE OF INVENTION: BIOCOMPATIBLE IMPLANT FOR THE EXPRESSION AND IN VIVO
; TITLE OF INVENTION: SECRETION OF A THERAPEUTIC SUBSTANCE
; FILE REFERENCE: 0660-0145-0DIV
; CURRENT APPLICATION NUMBER: US/09/225,509
; CURRENT FILING DATE: 1999-01-06
; EARLIER APPLICATION NUMBER: PCT/FR94/00456
; EARLIER FILING DATE: 1994-04-21
; EARLIER APPLICATION NUMBER: 09/523,814
; EARLIER FILING DATE: 1996-01-19
; EARLIER APPLICATION NUMBER: FR 93/04700
; EARLIER FILING DATE: 1993-04-21
; EARLIER APPLICATION NUMBER: FR 93/09185
; EARLIER FILING DATE: 1993-07-26
; NUMBER OF SEQ ID NOS: 1
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 8388
; TYPE: DNA

```

```

RESULT 11
US-09-225-509-1
; Sequence 1, Application US/09225509
; Patent No. 6326195
; GENERAL INFORMATION:
; APPLICANT: MOULLIER, Philippe
; APPLICANT: DANOS, Olivier
; APPLICANT: HEARD, Jean-Michel
; APPLICANT: FERRY, Nicholas
; TITLE OF INVENTION: BIOCOMPATIBLE IMPLANT FOR THE EXPRESSION AND IN VIVO
; TITLE OF INVENTION: SECRETION OF A THERAPEUTIC SUBSTANCE
; FILE REFERENCE: 0660-0145-0DIV
; CURRENT APPLICATION NUMBER: US/09/225,509
; CURRENT FILING DATE: 1999-01-06
; EARLIER APPLICATION NUMBER: PCT/FR94/00456
; EARLIER FILING DATE: 1994-04-21
; EARLIER APPLICATION NUMBER: 09/523,814
; EARLIER FILING DATE: 1996-01-19
; EARLIER APPLICATION NUMBER: FR 93/04700
; EARLIER FILING DATE: 1993-04-21
; EARLIER APPLICATION NUMBER: FR 93/09185
; EARLIER FILING DATE: 1993-07-26
; NUMBER OF SEQ ID NOS: 1
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 8388
; TYPE: DNA

```



```
; ORGANISM: mus musculus, Mo-MuLV, and other
US-09-225-509-1
Query Match      72.0%; Score 18; DB 4; Length 8388;
Best Local Similarity 100.0%; Pred. No. 16;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 8 GTGCAGGACGTGACAAAT 25
   |||||
Db 2238 GTGCAGGACGTGACAAAT 2255

RESULT 12
US-09-479-122-22
; Sequence 22, Application US/09479122
; Patent No. 6410266
; GENERAL INFORMATION:
; APPLICANT: HARRINGTON, JOHN J.
; APPLICANT: SHERF, BRUCE
; APPLICANT: RUNDLETT, STEPHEN
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR NON-TARGETED ACTIVATION OF
; FILE REFERENCE: 0221-0003C
; CURRENT APPLICATION NUMBER: US/09/479,122
; CURRENT FILING DATE: 2000-01-07
; PRIOR APPLICATION NUMBER: 09/276,820
; PRIOR FILING DATE: 1999-03-26
; PRIOR APPLICATION NUMBER: 09/159,643
; PRIOR FILING DATE: 1998-09-24
; PRIOR APPLICATION NUMBER: 08/941,223
; PRIOR FILING DATE: 1997-09-26
; PRIOR APPLICATION NUMBER: 09/263,814
; PRIOR FILING DATE: 1999-03-08
; PRIOR APPLICATION NUMBER: 09/253,022
; PRIOR FILING DATE: 1999-02-19
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 22
; LENGTH: 9737
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (8347)
; OTHER INFORMATION: a, c, t, g, other or unknown
; NAME/KEY: modified_base
; LOCATION: (8499)
; OTHER INFORMATION: a, c, t, g, other or unknown
US-09-479-122-23
Query Match      72.0%; Score 18; DB 4; Length 9737;
Best Local Similarity 100.0%; Pred. No. 17;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 8 GTGCAGGACGTGACAAAT 25
   |||||
Db 8356 GTGCAGGACGTGACAAAT 8373

RESULT 14
US-09-479-122-28
; Sequence 28, Application US/09479122
; Patent No. 6410266
; GENERAL INFORMATION:
; APPLICANT: HARRINGTON, JOHN J.
; APPLICANT: SHERF, BRUCE
; APPLICANT: RUNDLETT, STEPHEN
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR NON-TARGETED ACTIVATION OF
; FILE REFERENCE: 0221-0003C
; CURRENT APPLICATION NUMBER: US/09/479,122
; CURRENT FILING DATE: 2000-01-07
; PRIOR APPLICATION NUMBER: 09/276,820
; PRIOR FILING DATE: 1999-03-26
; PRIOR APPLICATION NUMBER: 09/159,643
; PRIOR FILING DATE: 1998-09-24
; PRIOR APPLICATION NUMBER: 08/941,223
; PRIOR FILING DATE: 1997-09-26
; PRIOR APPLICATION NUMBER: 09/263,814
; PRIOR FILING DATE: 1999-03-08
; PRIOR APPLICATION NUMBER: 09/253,022
; PRIOR FILING DATE: 1999-02-19
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 28
; LENGTH: 9737
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (8347)
; OTHER INFORMATION: a, c, t, g, other or unknown
; NAME/KEY: modified_base
; LOCATION: (8499)
; OTHER INFORMATION: a, c, t, g, other or unknown
US-09-479-122-28
Query Match      72.0%; Score 18; DB 4; Length 9737;
Best Local Similarity 100.0%; Pred. No. 17;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 8 GTGCAGGACGTGACAAAT 25
   |||||
Db 8356 GTGCAGGACGTGACAAAT 8373

RESULT 13
US-09-479-122-23
; Sequence 23, Application US/09479122
; Patent No. 6410266
; GENERAL INFORMATION:
; APPLICANT: HARRINGTON, JOHN J.
; APPLICANT: SHERF, BRUCE
; APPLICANT: RUNDLETT, STEPHEN
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR NON-TARGETED ACTIVATION OF
; FILE REFERENCE: 0221-0003C
; CURRENT APPLICATION NUMBER: US/09/479,122
; CURRENT FILING DATE: 2000-01-07
; PRIOR APPLICATION NUMBER: 09/276,820
```

Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 GTGCAGGACGTGACAAAT 25
 |||||
 Db 8356 GTGCAGGACGTGACAAAT 8373

RESULT 15
 US-09-479-122-24
 ; Sequence 24, Application US/09479122
 ; Patent No. 6410266
 ; GENERAL INFORMATION:
 ; APPLICANT: HARRINGTON, JOHN J.
 ; APPLICANT: SHEREF, BRUCE
 ; APPLICANT: RUNDLETT, STEPHEN
 ; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR NON-TARGETED ACTIVATION OF
 ; TITLE OF INVENTION: ENDOGENOUS GENES
 ; FILE REFERENCE: 0221-0003C
 ; CURRENT APPLICATION NUMBER: US/09/479,122
 ; CURRENT FILING DATE: 2000-01-07
 ; PRIOR APPLICATION NUMBER: 09/276,820
 ; PRIOR FILING DATE: 1999-03-26
 ; PRIOR APPLICATION NUMBER: 09/159,643
 ; PRIOR FILING DATE: 1998-09-24
 ; PRIOR APPLICATION NUMBER: 08/941,223
 ; PRIOR FILING DATE: 1997-09-26
 ; PRIOR APPLICATION NUMBER: 09/263,814
 ; PRIOR FILING DATE: 1999-03-08
 ; PRIOR APPLICATION NUMBER: 09/253,022
 ; PRIOR FILING DATE: 1999-02-19
 ; NUMBER OF SEQ ID NOS: 33
 ; SOFTWARE: PatentIn Ver. 2.1
 ; SEQ ID NO 24
 ; LENGTH: 9871
 ; TYPE: DNA
 ; ORGANISM: Homo sapiens
 ; FEATURE:
 ; NAME/KEY: modified_base
 ; LOCATION: (8481)
 ; OTHER INFORMATION: a, c, t, g, other or unknown
 ; NAME/KEY: modified_base
 ; LOCATION: (8633)
 ; OTHER INFORMATION: a, c, t, g, other or unknown
 US-09-479-122-24

Query Match 72.0%; Score 18; DB 4; Length 9871;
 Best Local Similarity 100.0%; Pred. No. 17;
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 GTGCAGGACGTGACAAAT 25
 |||||
 Db 8490 GTGCAGGACGTGACAAAT 8507

Search completed: January 3, 2003, 23:53:38
 Job time : 14.4385 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 3, 2003, 22:54:41 ; Search time 8.43849 Seconds
(without alignments)
1281.345 Million cell updates/sec

Title: US-09-787-562-1

Perfect score: 25

Sequence: 1 cgcctcggtgcaggacgtgacaaat 25

Scoring table: IDENTITY_NUC

Gapop 10.0, Gapext 1.0

Searched: 381593 seqs, 216252194 residues

Total number of hits satisfying chosen parameters: 763186

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Published Applications, NA.*
1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq.*
2: /cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq.*
3: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq.*
4: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq.*
5: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq.*
6: /cgn2_6/ptodata/1/pubpna/PCTUS_PUBCOMB.seq.*
7: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq.*
8: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq.*
9: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq.*
10: /cgn2_6/ptodata/1/pubpna/US09_PUBCOMB.seq.*
11: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq.*
12: /cgn2_6/ptodata/1/pubpna/US10_PUBCOMB.seq.*
13: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq.*
14: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	18	72.0	943	10	US-09-833-381-1259
2	18	72.0	4768	9	US-10-087-523-1
3	18	72.0	4768	10	US-09-816-790-1
4	18	72.0	4768	10	US-09-861-077-1
5	18	72.0	4768	10	US-09-815-825-1
6	18	72.0	4768	10	US-09-815-935-1
7	18	72.0	4768	10	US-09-815-944-1
8	18	72.0	6355	9	US-10-087-523-2
9	18	72.0	6355	10	US-09-816-790-2
10	18	72.0	6355	10	US-09-861-077-2
11	18	72.0	6355	10	US-09-815-825-2
12	18	72.0	6355	10	US-09-815-935-2
13	18	72.0	6355	10	US-09-815-944-2
14	18	72.0	8388	10	US-09-987-601-1
15	17	68.0	279	10	US-09-919-580-500
16	16.2	64.8	279	10	US-09-878-574-13168
17	16.2	64.8	538	10	US-09-998-598-458
18	16.2	64.8	729	9	US-09-738-626-1680
19	16.2	64.8	1797	9	US-09-738-626-1682

c	20	16	64.0	235	10	US-09-736-960-66	Sequence 66, Appl
c	21	16	64.0	534	9	US-09-738-626-2011	Sequence 2011, Ap
c	22	16	64.0	987	10	US-09-974-300-2046	Sequence 2046, Ap
c	23	16	64.0	1314	9	US-09-738-626-1619	Sequence 1619, Ap
c	24	16	64.0	1832	10	US-09-925-302-58	Sequence 58, Appl
c	25	16	64.0	3735	10	US-09-159-469-43	Sequence 43, Appl
c	26	16	64.0	3735	10	US-09-798-042-43	Sequence 43, Appl
c	27	16	64.0	3735	10	US-09-798-042-88	Sequence 88, Appl
c	28	16	64.0	3735	10	US-09-798-042-96	Sequence 96, Appl
c	29	16	64.0	4026	10	US-09-736-960-3	Sequence 3, Appl
c	30	16	64.0	7215	10	US-09-736-960-1	Sequence 1, Appl
c	31	16	64.0	43058	10	US-09-954-456-292	Sequence 292, App
c	32	16	64.0	43058	10	US-09-954-456-529	Sequence 529, App
c	33	16	64.0	43058	10	US-09-880-107-3950	Sequence 3950, Ap
c	34	16	64.0	66866	10	US-09-736-960-86	Sequence 86, Appl
c	35	16	64.0	3309400	9	US-09-738-626-1	Sequence 1, Appl
c	36	15.8	63.2	274	10	US-09-878-574-10968	Sequence 10968, A
c	37	15.8	63.2	461	10	US-09-974-300-8244	Sequence 8244, Ap
c	38	15.8	63.2	755	10	US-09-919-603-4	Sequence 4, Appl
c	39	15.8	63.2	925	10	US-09-919-603-6	Sequence 6, Appl
c	40	15.8	63.2	2439	9	US-09-954-531-140	Sequence 140, App
c	41	15.8	63.2	2439	9	US-09-954-531-359	Sequence 359, App
c	42	15.8	63.2	2748	10	US-09-822-849A-234	Sequence 234, App
c	43	15.8	63.2	43950	12	US-10-060-332-3	Sequence 3, Appl
c	44	15.6	62.4	274	9	US-09-996-634-72	Sequence 72, Appl
c	45	15.6	62.4	425	10	US-09-983-965-2423	Sequence 2423, Ap

ALIGNMENTS

RESULT 1

US-09-833-381-1259/c

; Sequence 1259, Application US/09833381

; Patent No. US20020132090A1

; GENERAL INFORMATION:

; APPLICANT: Robison, Keith E.

; TITLE OF INVENTION: No. US20020132090A1el Nucleic Acid and Protein Homologs

; FILE REFERENCE: 5800-119

; CURRENT APPLICATION NUMBER: US/09/833.381

; CURRENT FILING DATE: 2001-04-11

; PRIOR APPLICATION NUMBER: 09/516,448

; PRIOR FILING DATE: 2000-02-29

; NUMBER OF SEQ ID NOS: 2050

; SOFTWARE: FastSeq for Windows Version 3.0.

; SEQ ID NO 1259

; LENGTH: 943

; TYPE: DNA

; ORGANISM: Homo sapiens

; FEATURE:

; NAME/KEY: misc.feature

; LOCATION: (1)..(943)

; OTHER INFORMATION: n = A,T,C or G

US-09-833-381-1259

Query Match 72.0%; Score 18; DB 10; Length 943;

Best Local Similarity 100.0%; Pred. No. 7.1;

Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 8 GTGCAGGACGTGACAAAT 25

Db 153 GTGCAGGACGTGACAAAT 136

RESULT 2

US-10-087-523-1

; Sequence 1, Application US/10087523

; Publication No. US20020197624A1

; GENERAL INFORMATION:

; APPLICANT: Klein, Robert D.

; APPLICANT: Brennan, Thomas J.

; TITLE OF INVENTION: METHODS OF CREATING CONSTRUCTS USEFUL FOR INTRODUCING

; SEQUENCES INTO EMBRYONIC STEM CELLS

us-09-787-562-1.1.rnpb

Mon Jan 6 15:20:17 2003

; FILE REFERENCE: 376472000200
; CURRENT APPLICATION NUMBER: US/10/087,523
; CURRENT FILING DATE: 2002-02-28
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 09/193,834
; PRIOR FILING DATE: EARLIER FILING DATE: 1998-11-17
; NUMBER OF SEQ ID NOS: 44
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 1
; LENGTH: 4768
; TYPE: DNA
; ORGANISM: Plasmid vector
US-10-087-523-1

Query Match 72.0%; Score 18; DB 9; Length 4768;
Best Local Similarity 100.0%; Pred. No. 9;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 GTGCAGGACGTGACAAAT 25
|||||
Db 2824 GTGCAGGACGTGACAAAT 2841

RESULT 3
US-09-816-790-1
; Sequence 1, Application US/09816790
; Patent No. US2002002225A1
; GENERAL INFORMATION:
; APPLICANT: Allen, Keith D.
; APPLICANT: Phillips, Russell
; TITLE OF INVENTION: TRANSGENIC MICE CONTAINING
; TITLE OF INVENTION: SULFOTRANSFERASE GENE DISRUPTIONS
; FILE REFERENCE: R-855
; CURRENT APPLICATION NUMBER: US/09/816,790
; PRIOR FILING DATE: 2001-03-22
; PRIOR APPLICATION NUMBER: US 60/191,240
; PRIOR FILING DATE: 2000-03-22
; PRIOR APPLICATION NUMBER: US 60/204,230
; PRIOR FILING DATE: 2000-05-15
; PRIOR APPLICATION NUMBER: US 60/223,173
; PRIOR FILING DATE: 2000-08-07
; NUMBER OF SEQ ID NOS: 21
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 4768
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Phage vector
US-09-816-790-1

Query Match 72.0%; Score 18; DB 10; Length 4768;
Best Local Similarity 100.0%; Pred. No. 9;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 GTGCAGGACGTGACAAAT 25
|||||
Db 2824 GTGCAGGACGTGACAAAT 2841

RESULT 4
US-09-861-077-1
; Sequence 1, Application US/09861077
; Patent No. US20020023275A1
; GENERAL INFORMATION:
; APPLICANT: Leviten, Michael W.
; TITLE OF INVENTION: TRANSGENIC MICE CONTAINING MATRIX
; TITLE OF INVENTION: METALLOPROTEASE GENE DISRUPTIONS
; FILE REFERENCE: R-15
; CURRENT APPLICATION NUMBER: US/09/861,077
; PRIOR FILING DATE: 2000-05-17
; PRIOR APPLICATION NUMBER: US 60/204,972
; PRIOR FILING DATE: 2000-05-17
; PRIOR APPLICATION NUMBER: US 60/215,394

; PRIOR FILING DATE: 2000-06-29
; NUMBER OF SEQ ID NOS: 21
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 4768
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Phage vector
US-09-861-077-1

Query Match 72.0%; Score 18; DB 10; Length 4768;
Best Local Similarity 100.0%; Pred. No. 9;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 GTGCAGGACGTGACAAAT 25
|||||
Db 2824 GTGCAGGACGTGACAAAT 2841

RESULT 5
US-09-815-825-1
; Sequence 1, Application US/09815825
; Patent No. US20020026652A1
; GENERAL INFORMATION:
; APPLICANT: Allen, Keith D.
; APPLICANT: Phillips, Russell
; TITLE OF INVENTION: TRANSGENIC MICE CONTAINING CGMP
; TITLE OF INVENTION: PHOSPHODIESTERASE GENE DISRUPTIONS
; FILE REFERENCE: R-849
; CURRENT APPLICATION NUMBER: US/09/815,825
; PRIOR FILING DATE: 2001-03-22
; PRIOR APPLICATION NUMBER: US 60/191,142
; PRIOR FILING DATE: 2000-03-22
; PRIOR APPLICATION NUMBER: US 60/204,227
; PRIOR FILING DATE: 2000-05-15
; PRIOR APPLICATION NUMBER: US 60/216,765
; PRIOR FILING DATE: 2000-07-06
; PRIOR APPLICATION NUMBER: US 60/219,182
; PRIOR FILING DATE: 2000-07-19
; NUMBER OF SEQ ID NOS: 21
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 4768
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Phage vector
US-09-815-825-1

Query Match 72.0%; Score 18; DB 10; Length 4768;
Best Local Similarity 100.0%; Pred. No. 9;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 GTGCAGGACGTGACAAAT 25
|||||
Db 2824 GTGCAGGACGTGACAAAT 2841

RESULT 6
US-09-815-935-1
; Sequence 1, Application US/09815935
; Patent No. US20020038466A1
; GENERAL INFORMATION:
; APPLICANT: Allen, Keith D.
; TITLE OF INVENTION: TRANSGENIC MICE CONTAINING MAGNESIUM
; TITLE OF INVENTION: DEPENDENT PROTEIN PHOSPHATASE GENE DISRUPTIONS
; FILE REFERENCE: R-723
; CURRENT APPLICATION NUMBER: US/09/815,935
; PRIOR FILING DATE: 2001-03-22
; PRIOR APPLICATION NUMBER: US 60/191,235
; PRIOR FILING DATE: 2000-03-22
; PRIOR APPLICATION NUMBER: US 60/216,249

; PRIOR FILING DATE: 2000-07-06
; NUMBER OF SEQ ID NOS: 21
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 4768
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Phage vector
US-09-815-935-1

Query Match 72.0%; Score 18; DB 10; Length 4768;
Best Local Similarity 100.0%; Pred. No. 9;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 8 GTGCAGGACGTGACAAAT 25
|||||
Db 2824 GTGCAGGACGTGACAAAT 2841

RESULT 7
US-09-815-944-1
; Sequence 1, Application US/09815944
; Patent No. US20020038467A1
; GENERAL INFORMATION:
; APPLICANT: Allen, Keith D.
; APPLICANT: Matthews, William
; APPLICANT: Moore, Mark
; APPLICANT: Phillips, Russell
; TITLE OF INVENTION: TRANSGENIC MICE CONTAINING MELANOCYTE
; FILE REFERENCE: R-654
; CURRENT APPLICATION NUMBER: US/09/815,944
; CURRENT FILING DATE: 2001-03-22
; PRIOR APPLICATION NUMBER: US 60/191,236
; PRIOR FILING DATE: 2000-03-22
; PRIOR APPLICATION NUMBER: US 60/215,214
; PRIOR FILING DATE: 2000-06-29
; PRIOR APPLICATION NUMBER: US 60/218,075
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/219,167
; PRIOR FILING DATE: 2000-07-19
; NUMBER OF SEQ ID NOS: 21
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 4768
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Phage vector
US-09-815-944-1

Query Match 72.0%; Score 18; DB 10; Length 4768;
Best Local Similarity 100.0%; Pred. No. 9;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 8 GTGCAGGACGTGACAAAT 25
|||||
Db 2824 GTGCAGGACGTGACAAAT 2841

RESULT 8
US-10-087-523-2
; Sequence 2, Application US/10087523
; Publication No. US20020197624A1
; GENERAL INFORMATION:
; APPLICANT: Klein, Robert D.
; APPLICANT: Brennan, Thomas J.
; TITLE OF INVENTION: METHODS OF CREATING CONSTRUCTS USEFUL FOR INTRODUCING
; FILE REFERENCE: 376472000200
; CURRENT APPLICATION NUMBER: US/10/087,523
; CURRENT FILING DATE: 2002-02-28

; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 09/193,834
; PRIOR FILING DATE: EARLIER FILING DATE: 1998-11-17
; NUMBER OF SEQ ID NOS: 44
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 6355
; TYPE: DNA
; ORGANISM: Plasmid vector
US-10-087-523-2

Query Match 72.0%; Score 18; DB 9; Length 6355;
Best Local Similarity 100.0%; Pred. No. 9.4;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 8 GTGCAGGACGTGACAAAT 25
|||||
Db 4411 GTGCAGGACGTGACAAAT 4428

RESULT 9
US-09-816-790-2
; Sequence 2, Application US/09816790
; Patent No. US20020022255A1
; GENERAL INFORMATION:
; APPLICANT: Allen, Keith D.
; APPLICANT: Phillips, Russell
; TITLE OF INVENTION: TRANSGENIC MICE CONTAINING
; FILE REFERENCE: R-855
; CURRENT APPLICATION NUMBER: US/09/816,790
; CURRENT FILING DATE: 2001-03-22
; PRIOR APPLICATION NUMBER: US 60/191,240
; PRIOR FILING DATE: 2000-03-22
; PRIOR APPLICATION NUMBER: US 60/204,230
; PRIOR FILING DATE: 2000-05-15
; PRIOR APPLICATION NUMBER: US 60/223,173
; PRIOR FILING DATE: 2000-08-07
; NUMBER OF SEQ ID NOS: 21
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 6355
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Phage vector
US-09-816-790-2

Query Match 72.0%; Score 18; DB 10; Length 6355;
Best Local Similarity 100.0%; Pred. No. 9.4;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 8 GTGCAGGACGTGACAAAT 25
|||||
Db 4411 GTGCAGGACGTGACAAAT 4428

RESULT 10
US-09-861-077-2
; Sequence 2, Application US/09861077
; Patent No. US20020023275A1
; GENERAL INFORMATION:
; APPLICANT: Leviten, Michael W.
; TITLE OF INVENTION: TRANSGENIC MICE CONTAINING MATRIX
; FILE REFERENCE: R-15
; CURRENT APPLICATION NUMBER: US/09/861,077
; CURRENT FILING DATE: 2000-05-17
; PRIOR APPLICATION NUMBER: US 60/204,972
; PRIOR FILING DATE: 2000-05-17
; PRIOR APPLICATION NUMBER: US 60/215,394
; PRIOR FILING DATE: 2000-06-29
; NUMBER OF SEQ ID NOS: 21
; SOFTWARE: FastSeq for Windows Version 4.0

```
; SEQ ID NO 2
; LENGTH: 6355
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Phage vector
US-09-861-077-2

Query Match          72.0%; Score 18; DB 10; Length 6355;
Best Local Similarity 100.0%; Pred. No. 9.4;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 GTGCAGGACGTGACAAAT 25
   |||||
Db 4411 GTGCAGGACGTGACAAAT 4428

RESULT 11
US-09-815-825-2
; Sequence 2, Application US/09815825
; Patent No. US20020026652A1
; GENERAL INFORMATION:
; APPLICANT: Allen, Keith D.
; APPLICANT: Phillips, Russell
; TITLE OF INVENTION: TRANSGENIC MICE CONTAINING cGMP
; TITLE OF INVENTION: PHOSPHODIESTERASE GENE DISRUPTIONS
; FILE REFERENCE: R-849
; CURRENT APPLICATION NUMBER: US/09/815,825
; CURRENT FILING DATE: 2001-03-22
; PRIOR APPLICATION NUMBER: US 60/191,142
; PRIOR FILING DATE: 2000-03-22
; PRIOR APPLICATION NUMBER: US 60/204,227
; PRIOR FILING DATE: 2000-05-15
; PRIOR APPLICATION NUMBER: US 60/216,765
; PRIOR FILING DATE: 2000-07-06
; PRIOR APPLICATION NUMBER: US 60/219,182
; PRIOR FILING DATE: 2000-07-19
; NUMBER OF SEQ ID NOS: 21
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 6355
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Phage vector
US-09-815-825-2

Query Match          72.0%; Score 18; DB 10; Length 6355;
Best Local Similarity 100.0%; Pred. No. 9.4;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 GTGCAGGACGTGACAAAT 25
   |||||
Db 4411 GTGCAGGACGTGACAAAT 4428

RESULT 12
US-09-815-935-2
; Sequence 2, Application US/09815935
; Patent No. US20020038466A1
; GENERAL INFORMATION:
; APPLICANT: Allen, Keith D.
; TITLE OF INVENTION: TRANSGENIC MICE CONTAINING MAGNESIUM
; TITLE OF INVENTION: DEPENDENT PROTEIN PHOSPHATASE GENE DISRUPTIONS
; FILE REFERENCE: R-723
; CURRENT APPLICATION NUMBER: US/09/815,935
; CURRENT FILING DATE: 2001-03-22
; PRIOR APPLICATION NUMBER: US 60/191,235
; PRIOR FILING DATE: 2000-03-22
; PRIOR APPLICATION NUMBER: US 60/216,249
; PRIOR FILING DATE: 2000-07-06
; NUMBER OF SEQ ID NOS: 21
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 6355
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Phage vector
US-09-815-935-2

Query Match          72.0%; Score 18; DB 10; Length 6355;
Best Local Similarity 100.0%; Pred. No. 9.4;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 GTGCAGGACGTGACAAAT 25
   |||||
Db 4411 GTGCAGGACGTGACAAAT 4428

RESULT 13
US-09-815-944-2
; Sequence 2, Application US/09815944
; Patent No. US20020038467A1
; GENERAL INFORMATION:
; APPLICANT: Allen, Keith D.
; APPLICANT: Matthews, William
; APPLICANT: Moore, Mark
; APPLICANT: Phillips, Russell
; TITLE OF INVENTION: TRANSGENIC MICE CONTAINING MELANOCYTE
; TITLE OF INVENTION: STIMULATING HORMONE RECEPTOR GENE DISRUPTIONS
; FILE REFERENCE: R-654
; CURRENT APPLICATION NUMBER: US/09/815,944
; CURRENT FILING DATE: 2001-03-22
; PRIOR APPLICATION NUMBER: US 60/191,236
; PRIOR FILING DATE: 2000-03-22
; PRIOR APPLICATION NUMBER: US 60/215,214
; PRIOR FILING DATE: 2000-06-29
; PRIOR APPLICATION NUMBER: US 60/218,075
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/219,167
; PRIOR FILING DATE: 2000-07-19
; NUMBER OF SEQ ID NOS: 21
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 6355
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Phage vector
US-09-815-944-2

Query Match          72.0%; Score 18; DB 10; Length 6355;
Best Local Similarity 100.0%; Pred. No. 9.4;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 GTGCAGGACGTGACAAAT 25
   |||||
Db 4411 GTGCAGGACGTGACAAAT 4428

RESULT 14
US-09-987-601-1
; Sequence 1, Application US/09987601
; Patent No. US20020098223A1
; GENERAL INFORMATION:
; APPLICANT: MOULLIER, Philippe
; APPLICANT: DANOS, Olivier
; APPLICANT: HEARD, Jean-Michel
; APPLICANT: FERRY, Nicholas
; TITLE OF INVENTION: BIOCOMPATIBLE IMPLANT FOR THE EXPRESSION AND IN VIVO
; TITLE OF INVENTION: SECRETION OF A THERAPEUTIC SUBSTANCE
; FILE REFERENCE: 0660-0145-ODIV
; CURRENT APPLICATION NUMBER: US/09/987,601
; CURRENT FILING DATE: 2001-11-15
; PRIOR APPLICATION NUMBER: 2001-11-15
; PRIOR APPLICATION NUMBER: 09/225,509
```

Search completed: January 4, 2003, 01:06:05
Job time : 12.4385 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 3, 2003, 22:52:46 ; Search time 171.12 Seconds
(without alignments)
3231.380 Million cell updates/sec

Title: US-09-787-562-2

Perfect score: 19
Sequence: 1 gtcgtgcagagcgtgaca 19

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

GenEmbl:*

1: gb.ba:*
2: gb.htg:*
3: gb.in:*
4: gb.om:*
5: gb.ov:*
6: gb.pat:*
7: gb.ph:*
8: gb.pl:*
9: gb.pr:*
10: gb.ro:*
11: gb.sts:*
12: gb.sy:*
13: gb.un:*
14: gb.vi:*
15: em.ba:*
16: em.fun:*
17: em.hum:*
18: em.in:*
19: em.mu:*
20: em.om:*
21: em.or:*
22: em.ov:*
23: em.pat:*
24: em.ph:*
25: em.pl:*
26: em.ro:*
27: em.sts:*
28: em.un:*
29: em.vi:*
30: em.htg_hum:*
31: em.htg_inv:*
32: em.htg_other:*
33: em.htg_mus:*
34: em.htg_pln:*
35: em.htg_rtd:*
36: em.htg_mam:*
37: em.htg_vrt:*
38: em.sy:*
39: em.htgo_hum:*
40: em.htgo_mus:*
41: em.htgo_other:*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Match	Query Length	DB ID	Description
1	19	100.0	19	6	AX023660 Sequence
2	19	100.0	25	6	AX023659 Sequence
3	16.4	86.3	252715	2	AC127371 Mus muscu
4	16.4	86.3	338579	1	AP003004 Mesorhizo
5	16	84.2	1652	9	BC029804 Homo sapi
6	16	84.2	1777	6	AX463539 Sequence
7	16	84.2	50422	9	AL607085 Human DNA
8	16	84.2	118270	2	AL128395 Rattus no
9	16	84.2	121563	2	CNS07VQ4
10	16	84.2	200839	9	AC020763
11	15.8	83.2	461	6	AX439829
12	15.8	83.2	698	8	PU080658
13	15.8	83.2	699	8	CRU80647
14	15.8	83.2	807	8	PU080657
15	15.8	83.2	812	8	PU080660
16	15.8	83.2	826	8	AF323739
17	15.8	83.2	827	8	PU080656
18	15.8	83.2	828	8	PU080653
19	15.8	83.2	828	8	PU080661
20	15.8	83.2	869	8	AF287880
21	15.8	83.2	870	8	KSP406489
22	15.8	83.2	1052	8	AF357071
23	15.8	83.2	1228	9	HSU66579
24	15.8	83.2	1545	6	AX078628
25	15.8	83.2	1815	6	AR097359
26	15.8	83.2	19158	8	AF188714 Emericell
27	15.8	83.2	58439	2	AC100283
28	15.8	83.2	71596	2	AC100877
29	15.8	83.2	87022	2	AP001089
30	15.8	83.2	107336	2	AC126578
31	15.8	83.2	110000	2	AC009579_3
32	15.8	83.2	110000	2	AC124444_3
33	15.8	83.2	125515	2	AC100803
34	15.8	83.2	129302	2	AC083918
35	15.8	83.2	142362	2	AC036162
36	15.8	83.2	143239	2	AP005285
37	15.8	83.2	154926	9	AP005263
38	15.8	83.2	159272	2	AC131024
39	15.8	83.2	161674	2	AC015567
40	15.8	83.2	165591	9	AC019239
41	15.8	83.2	169124	2	AP001375
42	15.8	83.2	173368	9	AC091647
43	15.8	83.2	183030	2	AC130772
44	15.8	83.2	185479	2	AP001484
45	15.8	83.2	186050	1	AL646072 Ralstonia

ALIGNMENTS

RESULT 1	AX023660	Sequence 2	19 bp	DNA	linear	PAT 15-SEP-2000
AX023660	Sequence 2	from Patent WO0017371.				
LOCUS	AX023660					
DEFINITION	Sequence 2	from Patent WO0017371.				
ACCESSION	AX023660					
VERSION	AX023660.1	GI:10184021				
KEYWORDS						
SOURCE	Mus sp.					
ORGANISM	Mus sp.					
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;					
AUTHORS	1 (bases 1 to 19)					
TITLE	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.					
JOURNAL	Binley, K.M. and Naylor, S.					
	Polynucleotide constructs and uses thereof					
	Patent: WO 0017371-A 2 30-MAR-2000;					

McPherson, J.D. and Waterston, R.H.
 Direct Submission
 Submitted (17-JUL-2002) Genome Sequencing Center, 4444 Forest Park
 Parkway, St. Louis, MO 63108, USA

----- Genome Center -----
 Center: Washington University Genome Sequencing Center
 Center code: WUGSC
 Web site: http://genome.wustl.edu/gsc/index.shtml
 Contact: submissions@wustl.edu
 Project information -----
 Center project name: M_BA0035D04

 ----- Summary Statistics -----
 Sequencing vector: M13; 0%
 Sequencing vector: plasmid; 100%
 Chemistry: Dye-terminator Big Dye; 100% of reads
 Assembly program: Phrap; version 0.990319
 Consensus quality: 246396 bases at least Q40
 Consensus quality: 246893 bases at least Q30
 Consensus quality: 247291 bases at least Q20
 Insert size: 218000; agarose-fp
 Insert size: 251715; sum-of-contigs
 Quality coverage: 0.00 in Q20 bases; agarose-fp
 Quality coverage: 11.44 in Q20 bases; sum-of-contigs

 * NOTE: This is a 'working draft' sequence. It currently
 * consists of 11 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

1 1247: contig of 1247 bp in length
 1248: gap of unknown length
 1348: contig of 2449 bp in length
 3797: gap of unknown length
 3897: contig of 3042 bp in length
 6938: gap of unknown length
 7038: contig of 16260 bp in length
 23298: gap of unknown length
 23398: contig of 16042 bp in length
 39440: gap of unknown length
 39541: contig of 23707 bp in length
 63247: gap of unknown length
 63348: contig of 45158 bp in length
 108505: gap of unknown length
 108605: contig of 31361 bp in length
 139966: gap of unknown length
 140066: contig of 109286 bp in length
 140067: gap of unknown length
 249352: contig of 1547 bp in length
 249353: gap of unknown length
 250999: contig of 1547 bp in length
 251099: gap of unknown length
 251100: contig of 1616 bp in length.

Location/Qualifiers
 1..252715
 /organism="Mus musculus"
 /db_xref="taxon:10090"
 /chromosome="UNK"
 /clone="RP23-35D4"
 1..1247
 /note="assembly_name:Contig10"
 1348..3796
 /note="assembly_name:Contig11"
 3897..6938
 /note="assembly_name:Contig12"
 7039..23298
 /note="assembly_name:Contig13"
 23399..39440
 /note="assembly_name:Contig14"
 39541..63247

misc_feature
 misc_feature
 misc_feature
 misc_feature
 misc_feature
 misc_feature

FEATURES
 source

1 19
 /organism="Mus sp."
 /db_xref="taxon:10095"
 4 a 4 c 8 g 3 t

Query Match 100.0%; Score 19; DB 6; Length 19;
 Best Local Similarity 100.0%; Pred. No. 3.1e+02;
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GTCGGTGCAGGACGTGACA 19
 |||||
 Db 1 GTCGGTGCAGGACGTGACA 19

RESULT 2
 AX023659
 LOCUS
 DEFINITION Sequence 1 from Patent WO0017371.
 ACCESSION AX023659
 VERSION AX023659.1 GI:10184020
 KEYWORDS
 SOURCE Mus sp.
 ORGANISM Mus sp.
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 25)
 Binley, K.M. and Naylor, S.
 Polynucleotide constructs and uses thereof
 Patent: WO 0017371-A 1 30-MAR-2000;
 JOURNAL BINLEY KATIE MARY (GB) ; NAYLOR STUART (GB) ; OXFORD BIOMEDICA LTD (GB)

FEATURES
 source
 1..25
 /organism="Mus sp."
 /db_xref="taxon:10095"
 6 a 6 c 9 g 4 t

Query Match 100.0%; Score 19; DB 6; Length 25;
 Best Local Similarity 100.0%; Pred. No. 3.1e+02;
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GTCGGTGCAGGACGTGACA 19
 |||||
 Db 4 GTCGGTGCAGGACGTGACA 22

RESULT 3
 AC127371
 LOCUS
 DEFINITION Mus musculus chromosome UNK clone RP23-35D4, WORKING DRAFT
 SEQUENCE, 11 unordered pieces.
 AC127371
 AC127371.1 GI:21747785
 VERSION HTG; HTGS_PHADEL; HTGS_DRAFT.
 KEYWORDS house mouse.
 SOURCE Mus musculus
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 252715)
 McPherson, J.D. and Waterston, R.H.
 The sequence of Mus musculus clone
 Unpublished
 JOURNAL McPherson, J.D. and Waterston, R.H.
 2 (bases 1 to 252715)
 McPherson, J.D. and Waterston, R.H.
 Direct Submission
 Submitted (14-JUL-2002) Genome Sequencing Center, 4444 Forest Park
 Parkway, St. Louis, MO 63108, USA
 3 (bases 1 to 252715)

REFERENCE
 AUTHORS
 TITLE
 JOURNAL
 REFERENCE
 AUTHORS
 TITLE
 JOURNAL
 REFERENCE

AUTHORS
 TITLE
 JOURNAL
 COMMENT

ene

/gene="mlr4362"
 /note="hypothetical protein"
 /codon_start=1
 /transl_table=11
 /protein_id="BAB51036.1"
 /db_xref="GI:14024433"
 /translation="MTRLVAYLVRMLFGLGVAVASLASAFLNVLFLASLGYPDPHT
 LPTATASIFSPFALFVAFPAFVILVAILGRDMLXYALAGAVVATVFLG
 LVDAHQESTFAITDSAINAVIGGVMGFIYWLISAGRWAGSWKDEASSTSPGPGS
 "

gene
 CDS
 complement(5769..6410)
 /gene="mlr4363"
 complement(5769..6410)
 /gene="mlr4363"
 /note="hypothetical protein"
 /codon_start=1
 /transl_table=11
 /protein_id="BAB51037.1"
 /db_xref="GI:14024434"
 /translation="MSFIPDTATLQFAIATVILATPGPDMTLFVSRTLSSGRATGF
 ASMAALCGTLIHTTLVVGVSALIVASMAFFVLKIFGAGYLVLAWOAIKAGSAFS
 PEKKTGPOISLLRSWAGLVNLPKIIILFFMTFLPQVSAHDPAKGLFFLGLMF
 IVLSIPVTAPMVLAAERFSAAMKASPRVTRVVDYLFAGVFSFALKILTAQAK"
 complement(6530..8746)
 /gene="mlr4364"
 complement(6530..8746)
 /gene="mlr4364"
 /note="hypothetical protein"
 /codon_start=1
 /transl_table=11
 /protein_id="BAB51038.1"
 /db_xref="GI:14024435"
 /translation="MGCIIARDLSKFSRPAKATQNCQGHATVAKRLPARCGCLGL
 GPFMRVTSCTIETHNLWLALMCVTCGWTGIIGLDRARKTVGQMRGLFUTAV
 AAGSITWCTHIFAMLTQPGAPITFDPAFTMSLVIAVMGTGAGFLADKGRGLAPE
 WGGCLVGLIAVLTGHTAMAAVATPLSFTTGTNPDIEMAAVAVAVAGLIYAA
 WSHVVGILSLAIVLHFTAMAAVATPLSFTTGTNPDIEMAAVAVAVAGLIYAA
 TGFASYLDERGLESERLOHALNDSLTGLANVSFNDRLDHEIAREHEDQMTA
 VIVIDLRKEINDRLRHAAGDOAKLIIARLAKLGDGEFVARLGDDEFAALKRFD
 LNDLGLVRLKSLFLEPLRIDDFETVIGASIGVAVYPRGADRERLVSNADLAMYRA
 KNDTRAVCFYSAMDETARARALATDLROADRGELSUIHQVQTSVQGTGCGHEA
 LLRWTHPVHGMIPAEFTIAEENGSLIAGEWLRTACQAAASWDNGHKAIVNLSPV
 OFAHADLATVHLQVLETPISPKLELETESTIVADKVTLHLVLRQIKALGTIYID
 DQTCYSLSLDTLRSFPEDKIKLDRSEWADVERSPOAKIIRAVLTIGRSLIDIPVLAEG
 VETHVOLRILOVEGCNEAQGYFLGRPKPIDQIGLISGADTGAPTEFVPMVRVAIG"
 8951..10174
 /gene="mlr4366"
 8951..10174
 /gene="mlr4366"
 /codon_start=1
 /transl_table=11
 /product="argininosuccinate synthase"
 /protein_id="BAB51039.1"
 /db_xref="GI:14024436"
 /translation="MSKTKDKVKVVLAYVSGGLDTSIIKWIQELGAEVVTFTADLQ
 GGELEPARKAEMMGIKDIRIVDVREEFVDFVPMFRANTVYEGTYLLGTSIARPLI
 SKHLVDIARETCADAIAHGATGKNDQVRFELSAYALNPDIKVIAPWRDRSPKSTDL
 INFADHOIIPVAKDRGRAPESVDANLHSSSEKGLVEDPWPSEPEFHQRTVPMDA
 PKVTEIEFELKGPALNGKLLSPATMLAALNDIGRONGIGRLDLVENVRFYGMKSR
 GYETPPGAILIVHRAIESITLDGRGAHLKDEMPRYAELIYNGFWFSPERLMLQAM
 IDKQDEVEGTVRLKLYKGNVNTGRKSKTLYISDALVTFEDDRGATDQKDAAGFIFL
 NALRLFTLAARNRG"
 10349..10786
 /gene="mlr4368"
 10349..10786
 /gene="mlr4368"
 /note="unknown protein"
 /codon_start=1
 /transl_table=11
 /protein_id="BAB51040.1"
 /db_xref="GI:14024437"
 /translation="MKNSTSLILGAVVVLVSCSSQAGAKRDPKVDMECIASARTSN
 LTPKHDWASFMVGSQERNPALFCQRLADGIRTSYSDINNLQDQTEIWMVLKKG

PKRAAPTHSOAPRSPNFRNCSDGIDGAFQVAPASQRCPASGYAHY"
 10829..11314
 /gene="mlr4369"
 10829..11314
 /gene="mlr4369"
 /note="unknown protein"
 Query Match 86.3%; Score 16.4; DB 1; Length 338579;
 Best Local Similarity 94.4%; Pred. No. 2.2e+03;
 Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Oy 2 TCGGTGCAGGACGTGACA 19
 ||||| ||||| ||||| |||||
 Db 60050 TCGGTGCAGGACGTGACA 60067

RESULT 5
 BC029804
 LOCUS Homo sapiens, LOC146433, clone MGC:34647 IMAGE:5192904, mRNA, 1652 bp
 DEFINITION complete cds.
 ACCESSION BC029804
 VERSION BC029804.1 GI:20987449
 KEYWORDS MGC.
 SOURCE Homo sapiens.
 ORGANISM Homo sapiens
 Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (bases 1 to 1652)
 AUTHORS Strausberg, R.
 TITLE Direct Submission
 JOURNAL Submitted (06-MAY-2002) National Institutes of Health, Mammalian
 Gene Collection (MGC), Cancer Genomics Office, National Cancer
 Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
 USA
 NIH-MGC project URL: <http://mgc.nci.nih.gov>
 Contact: MGC help desk
 Email: cdapbs-r@mail.nih.gov
 Tissue Procurement: Life Technologies, Inc.
 cDNA Library Preparation: Life Technologies, Inc.
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Baylor College of Medicine Human Genome
 Sequencing Center
 Center code: BCM-HGSC
 Web site: <http://www.hgsc.bcm.tmc.edu/cdna/>
 Contact: amg@bcm.tmc.edu
 Gunaratne, P.H., Garcia, A.M., Lu, X., Hulyk, S.W., Hale, S.M.,
 Yoon, V.S., Kowis, C.R., Lawrence, S., Martin, R.G., Muzny, D.M.,
 Richards, S., Gibbs, R.A.

Clone distribution: MGC clone distribution information can be found
 through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
 Series: IRAK Plate: 50 Row: 1 Column: 18
 This clone was selected for full length sequencing because it
 passed the following selection criteria: Hexamer frequency ORF
 analysis, Similarity but not identity to protein.

FEATURES
 Location/Qualifiers
 1..1652
 /organism="Homo sapiens"
 /db_xref="LocusID:146433"
 /db_xref="taxon:9606"
 /clone="MGC:34647 IMAGE:5192904"
 /tissue_type="Brain, adult, 6 pooled whole brains"
 /clone_lib="NH_MGC_114"
 /lab_host="DH10B"
 /note="Vector: pCMV-SPORT6"
 384..1112
 /codon_start=1
 /product="LOC146433"
 /protein_id="BAB29804.1"
 /db_xref="GI:20987450"
 /translation="MPRGFTWIRYIGIFGLVALGNEPLEWPLTONSECTVTCFLRDK
 LQYRSRLQYMKHYFPYKISVPYEGVFRIANVTNLQRAQVSRRELRYLWVLSLAT

ESVDVLEHPSWKYLOEVOTLLNVOOGLTDEVEVSPKVESVLNADPNLKLVR
PKALDNCFRVWELLYCSCCKOSSVLNWDDEVSPSCSPSPSLQTAATQLIPPPPW
SPSPPHSTGSRVPVRAQGEGLIP*
BASE COUNT 336 a 504 c 499 g 313 t
ORIGIN

Query Match 84.2%; Score 16; DB 9; Length 1652;
Best Local Similarity 100.0%; Pred. No. 5.2e+03;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GTCGGTGACGACGCG 16
|||||
Db 692 GTCGGTGACGACGCG 707

RESULT 6
AX463539
LOCUS AX463539 1777 bp DNA linear PAT 15-JUL-2002
DEFINITION Sequence 103 from Patent WO0248337.
ACCESSION AX463539
VERSION AX463539.1 GI:21886313
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens

REFERENCE
AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
TITLE Mammalia; Euthera; Primates; Catarrhini; Hominiidae; Homo.
JOURNAL

SECRETED PROTEINS
Duggan, B.M., Yao, M.G. and Griffin, J.A.
Patent: WO 0248337-A 103 20-JUN-2002;
INCYTE GENOMICS INC (US)

FEATURES
Location/Qualifiers
source
1..1777
/organism="Homo sapiens"
/db_xref="taxon:9606"
/notes="Incyte ID No: 7486536CB1"
BASE COUNT 331 a 555 c 548 g 343 t
ORIGIN

Query Match 84.2%; Score 16; DB 6; Length 1777;
Best Local Similarity 100.0%; Pred. No. 5.2e+03;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GTCGGTGACGACGCG 16
|||||
Db 851 GTCGGTGACGACGCG 866

RESULT 7
AL607085/c
LOCUS AL607085 50422 bp DNA linear PRI 08-APR-2002
DEFINITION Human DNA sequence from clone Rpl1-529L18 on chromosome 10,
complete sequence.

ACCESSION AL607085
VERSION AL607085.8 GI:20135753
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Euthera; Primates; Catarrhini; Hominiidae; Homo.
1

Heath, P.
Submitted (08-APR-2002) Wellcome Trust Sanger Institute, Hinxton,
Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
humquery@sanger.ac.uk

COMMENT
On Apr 10, 2002 this sequence version replaced gi:16974012.
During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variation annotation may not be found in the sequence submission
corresponding to the overlapping clone, as we submit sequences with
only a small overlap as described above.

This sequence was finished as follows unless otherwise noted: all
regions were either double-stranded or sequenced with an alternate
chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such
as compressions and repeats; all regions were covered by at least
one plasmid subclone or more than one M13 subclone; and the
assembly was confirmed by restriction digest. The following
abbreviations are used to associate primary accession numbers given
in the feature table with their source databases: Em; EMBL; SWI;
SWISSPROT; Tr; TREMBL; Wp; WORMPEP; Information on the WORMPEP
database can be found at
http://www.sanger.ac.uk/projects/Celegans/wormpep This sequence
was generated from part of bacterial clone contigs of human
chromosome 10, constructed by the Sanger Centre Chromosome 10
Mapping Group. Further information can be found at
http://www.sanger.ac.uk/HGP/Chr10
Rpl1-529L18 is from the library RPl1-11.2 constructed by the group
of Pieter de Jong. For further details see
http://www.chori.org/bacpac/home.htm
VECTOR: pBACE3.6.

FEATURES
Location/Qualifiers
source
1..50422
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="10"
/clone="Rpl1-529L18"
/clone_lib="RPl1-11.2"

BASE COUNT 12195 a 11534 c 12203 g 14490 t
ORIGIN

Query Match 84.2%; Score 16; DB 9; Length 50422;
Best Local Similarity 100.0%; Pred. No. 4e+03;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 GGTGCAGGACGTGACA 19
|||||
Db 2520 GGTGCAGGACGTGACA 2505

RESULT 8
AC128395/c
LOCUS AC128395 118270 bp DNA linear HTG 19-JUL-2002
DEFINITION Rattus norvegicus clone CH230-128B22, *** SEQUENCING IN PROGRESS
***, 43 unordered pieces.

ACCESSION AC128395
VERSION AC128395.1 GI:21909054
KEYWORDS HTG; HTGS_PHASE1.
SOURCE Rattus norvegicus.
ORGANISM Rattus norvegicus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Euthera; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.

REFERENCE
1 (bases 1 to 118270)
AUTHORS Muzny, D.M., Adams, C., Adio-Oduola, B., Ali-osman, F.R., Allen, C.,
Alsbrooks, S.L., Amaratunge, H.C., Are, J.R., Ayele, M., Banks, T.,
Barbaria, J., Benton, J., Bimage, K., Blankenburg, K., Bonnin, D.,
Bouck, J., Bowie, S., Brieva, M., Brown, E., Brown, M., Bryant, N.P.,
Buhay, C., Burck, P., Burkett, C., Burrell, K.L., Byrd, N.C.,
Carron, T.F., Carter, M., Cavazos, S.R., Chacko, J., Chavez, D.,
Chen, G., Chen, R., Chen, Z., Chowdhry, I., Christopoulos, C.,
Cleaveland, C.D., Cox, C., Coyle, M.D., Dathorne, S.R., David, R.,
Davila, M.L., Davis, C., Davy-Carroll, L., Dederich, D.A.,
Delaney, K.R., Delgado, O., Denn, A.L., Ding, Y., Dinh, H.H.,
Douthwaite, K.J., Draper, H., Dugan-Rocha, S., Durbin, K.J.,
Earnhart, C., Edgar, D., Edwards, C.C., Elhaj, C., Escotto, M.,
Falls, T., Ferraguto, D., Flagg, N., Ford, J., Foster, P., Frantz, P.,
Gabis, A., Gao, J., Garcia, A., Garner, T., Garza, N., Gill, R.,
Gorrell, J.H., Guevara, W., Gunaratne, P., Hale, S., Hamilton, K.,
Harris, C., Harris, K., Hart, M., Havlak, P., Hawes, A., Hernandez, J.,
Hernandez, O., Hodson, A., Hogues, M., Holloway, C., Hollins, B.,
Homs, F., Howard, S., Huber, J., Hulyk, S., Hume, J., Jackson, L.E.,
Jacobson, B., Jia, X., Johnson, R., Jolivet, S., Joudah, S.,
Karlsson, E., Kelly, S., Khan, U., King, L., Korvah, J., Kovar, C.,

Kratovic, J., Kureshi, A., Landry, N., Leal, B., Lewis, L. C., Lewis, L., Li, J., Li, Z., Licharge, O., Liu, C., Liu, J., Liu, W., Loulsegue, H., Lozano, R. J., Lu, X., Lucier, A., Lucier, R., Luna, R., Ma, J., Maheshwari, M., Mapua, E., Martin, R., Martindale, A., Martinez, E., Massey, E., Mawhney, E., McLeod, M. P., Meador, M., Mei, G., Metzker, M., Miner, G., Miner, Z., Mitchell, T., Mohabbat, K., Morgan, M., Morris, S., Moser, M., Neal, D., Newton, J., Newton, N., Nguyen, A., Nguyen, N., Nguyen, N., Nickerson, E., Nwokenkwo, S., Oguh, M., Okwuonu, G., Oragunye, N., Oviedo, R., Pace, A., Payton, B., Peery, J., Perez, L., Peters, L., Pickens, R., Primus, E., Pu, L. L., Quiles, M., Ren, Y., Rives, M., Rojas, A., Rojibokan, I., Rolfe, M., Ruiz, S., Savary, G., Scherer, S., Scott, G., Shen, H., Shooshtari, N., Sisson, I., Sodergren, E., Sonaike, T., Sparks, A., Stanley, H., Stone, H., Sutton, A., Svatek, A., Tabor, P., Tamerisa, A., Tamerisa, K., Tang, H., Tansey, J., Taylor, C., Taylor, R., Telford, B., Thomas, N., Thomas, S., Usmani, K., Vasquez, L., Vera, V., Villalobos, D., Vinson, R., Wang, Q., Wang, S., Ward-Moore, S., Warren, R., Washington, C., Watlington, S., Williams, G., Williamson, A., Wleczyk, R., Wooden, S., Worley, K., Wu, C., Wu, Y., Wu, Y. F., Zhou, J., Zorrilla, S., Nelson, D., Weinstein, G., and Gibbs, R.

Direct Submission
Unpublished
2 (bases 1 to 118270)
Worley, K.C.

Direct Submission
Submitted (19-JUL-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: <http://www.hgsc.bcm.tmc.edu/>
Contact: hgsc-help@bcm.tmc.edu

----- Project Information
Center project name: GRZ1
Center clone name: CH230-128B22
----- Summary Statistics
Sequencing vector: Plasmid
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.990329
Consensus quality: 69006 bases at least Q40
Consensus quality: 71372 bases at least Q30
Consensus quality: 73129 bases at least Q20

* NOTE: Estimated insert size may differ from sequence length
(see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 43 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

1
1020: contig of 1020 bp in length
1021
1120: gap of unknown length
1121
2392: contig of 1172 bp in length
2293
2392: gap of unknown length
2293
3673: contig of 1281 bp in length
3674
3773: gap of unknown length
3774
5076: contig of 1303 bp in length
5077
5176: gap of unknown length
5177
6770: contig of 1594 bp in length
6771
6870: gap of unknown length
6871
8638
8737: gap of unknown length
8638
9864: contig of 1127 bp in length
9865
9965
10986: contig of 1022 bp in length
10987
11086: gap of unknown length
11087
12691: contig of 1605 bp in length
12692
12791: gap of unknown length
12792
14343: contig of 1552 bp in length

14344
14444
15721: contig of 1278 bp in length
15821: gap of unknown length
17281: contig of 1460 bp in length
17381: gap of unknown length
17382
18908: contig of 1527 bp in length
18909
19008: gap of unknown length
20601: contig of 1593 bp in length
20701: gap of unknown length
22259: contig of 1558 bp in length
22359: gap of unknown length
22358: contig of 1139 bp in length
23498: gap of unknown length
23598: contig of 1531 bp in length
25129: gap of unknown length
25229: contig of 1506 bp in length
26735: gap of unknown length
26835: gap of unknown length
28367: contig of 1532 bp in length
28467: gap of unknown length
29978: contig of 1511 bp in length
30078: gap of unknown length
31451: contig of 1373 bp in length
31551: gap of unknown length
33097: contig of 1546 bp in length
33197: gap of unknown length
34788: contig of 1591 bp in length
34888: gap of unknown length
36737: contig of 1849 bp in length
36837: gap of unknown length
38328: contig of 1491 bp in length
38428: gap of unknown length
40475: contig of 2047 bp in length
40476
40476: gap of unknown length
42115: contig of 1540 bp in length
42116
42116: gap of unknown length
43596: contig of 1381 bp in length
43597
43597: gap of unknown length
45346: contig of 1650 bp in length
45347
45347: gap of unknown length
47250: contig of 1804 bp in length
47251
47251: gap of unknown length
49576: contig of 2226 bp in length
49577
49577: gap of unknown length
52558: contig of 2882 bp in length
52559
52559: gap of unknown length
52658: gap of unknown length
54645: contig of 1987 bp in length
54646
54745: gap of unknown length
58870: contig of 4125 bp in length
58871
58970: gap of unknown length
63137: contig of 4167 bp in length
63237: gap of unknown length
68954: contig of 5717 bp in length
69054: gap of unknown length
74577: contig of 5523 bp in length
74578
74578: gap of unknown length
81209: contig of 6532 bp in length
81210
81309: gap of unknown length
86316: contig of 5007 bp in length
86317
86317: gap of unknown length
91067: contig of 4651 bp in length
91068
91167: gap of unknown length
100053: contig of 8886 bp in length
100054
100153: gap of unknown length
100154
107094: contig of 6941 bp in length
107095
107194: gap of unknown length
118270: contig of 11076 bp in length.

Location/Qualifiers
1. 118270
/organism="Rattus norvegicus"
/db_xref="taxon:10116"
/clone="CH230-128B22"

29719 a 22572 c 22749 g 29078 t 14152 others
BASE COUNT
ORIGIN

Query Match 84.2%; Score 16; DB 2; Length 118270;
 Best Local Similarity 100.0%; Pred. No. 3.7e+03;
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 TCGGTGCAGGACGTGA 17
 |||||

Db 48602 TCGGTGCAGGACGTGA 48587

RESULT 9
 CNS07YQ4/C
 LOCUS
 DEFINITION Oryza sativa chromosome 12 clone OJ1306_H03, *** SEQUENCING IN
 PROGRESS ***, in ordered pieces.
 ACCESSION AL713904
 VERSION AL713904.3 GI:20513132
 KEYWORDS HTG; HTGS; PHASE2; HTGS_ACTIVEFIN.
 SOURCE Oryza sativa.
 ORGANISM Oryza sativa

REFERENCE
 AUTHORS Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
 Ehrhartoideae; Oryzaceae; Oryza.
 1 (bases 1 to 121563)
 Choisne, N., Orjeda, G., Cattolico, L., Demange, N., Wincker, P.,
 Seguren, B., Pelletier, E., Scarpetti, C., Salanoubat, M.,
 Weissenbach, J., and Quetier, F.
 Oryza sativa chromosome 12 sequencing
 Unpublished
 2 (bases 1 to 121563)
 Genoscope.
 Direct Submission
 Submitted (06-MAY-2002) Genoscope - Centre National de Sequencage :
 BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
 - Web : www.genoscope.cns.fr)
 On May 8, 2002 this sequence version replaced gi:20160268.
 IMPORTANT: This sequence is unfinished and does not necessarily
 represent the correct sequence.
 Work on the sequence is in progress and the release of this data is
 based on the understanding that the sequence may change as work
 continue. The sequence may be contaminated with foreign sequence
 from E.coli, yeast, vector, phage, etc.
 The nucleotide sequence of this BAC clone was generated by
 combining Monsanto and Genoscope sequencing data.
 * NOTE: This is a 'working draft' sequence.
 * This sequence will be replaced
 * by the finished sequence as soon as it is available and
 * the accession number will be preserved.

FEATURES
 source
 Location/Qualifiers
 1..121563
 /organism="Oryza sativa"
 /cultivar="Nipponbare"
 /sub_species="japonica"
 /db_xref="taxon:4530"
 /chromosome="12"
 /clone="OJ1306_H03"
 /clone.lib="Monsanto"

BASE COUNT 33858 a 26851 c 27291 g 33563 t
 ORIGIN

Query Match 84.2%; Score 16; DB 2; Length 121563;
 Best Local Similarity 100.0%; Pred. No. 3.7e+03;
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GTCGGTGCAGGACGTG 16
 |||||

Db 71663 GTCGGTGCAGGACGTG 71648

RESULT 10
 AC020763/C
 LOCUS
 DEFINITION Homo sapiens chromosome 16 clone RP11-394B2, complete sequence.

ACCESSION AC020763
 VERSION AC020763.5 GI:22122869
 KEYWORDS HTG.
 SOURCE human.
 ORGANISM Homo sapiens

REFERENCE
 AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 200839)
 DOE Joint Genome Institute.
 TITLE Sequencing of Human Chromosome 16
 JOURNAL Unpublished
 REFERENCE 2 (bases 1 to 200839)
 DOE Joint Genome Institute.
 TITLE Direct Submission
 JOURNAL Submitted (09-JAN-2000) Production Sequencing Facility, DOE Joint
 Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
 REFERENCE 3 (bases 1 to 200839)
 DOE Joint Genome Institute.
 TITLE Direct Submission
 JOURNAL Submitted (06-AUG-2002) Production Sequencing Facility, DOE Joint
 Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
 COMMENT On Aug 6, 2002 this sequence version replaced gi:8576110.
 Sequence Quality Assessment:
 This entry has been annotated with sequence quality
 estimates computed by the Phrap assembly program.
 All manually edited bases have been reduced to quality zero.
 Quality levels above 40 are expected to have less than
 1 error in 10,000 bp.
 Base-by-base quality values are not generally visible from the
 GenBank flat file format but are available as part
 of this entry's ASN.1 file.

 Sequence Quality Assessment:
 This entry has been annotated with sequence quality
 estimates computed by the Phrap assembly program.
 All manually edited bases have been reduced to quality zero.
 Quality levels above 40 are expected to have less than
 1 error in 10,000 bp.
 Base-by-base quality values are not generally visible from the
 GenBank flat file format but are available as part
 of this entry's ASN.1 file.

 FEATURES
 Location/Qualifiers
 source
 1..200839
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="16"
 /clone="RP11-394B2"

BASE COUNT 47780 a 54805 c 50908 g 47346 t
 ORIGIN

Query Match 84.2%; Score 16; DB 9; Length 200839;
 Best Local Similarity 100.0%; Pred. No. 3.6e+03;
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GTCGGTGCAGGACGTG 16
 |||||

Db 116671 GTCGGTGCAGGACGTG 116656

RESULT 11
 AX439829
 LOCUS
 DEFINITION Sequence 8244 from Patent WO0229113.
 ACCESSION AX439829
 VERSION AX439829.1 GI:21664640
 KEYWORDS
 SOURCE Bacillus clausii.
 ORGANISM Bacillus clausii
 Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
 REFERENCE 1
 AUTHORS Berka, R. and Clausen, I.G.
 TITLE Methods for monitoring multiple gene expression

Mon Jan 6 15:20:22 2003

JOURNAL Patent: WO 0229113-A 8244 11-APR-2002;
Novozymes Biotech, Inc. (US); Novozymes A/S (DK)
FEATURES Location/Qualifiers
source 1..461
/organism="Bacillus clausii"
/db_xref="taxon:79880"

BASE COUNT 124 a 118 g 125 t
ORIGIN

Query Match 83.2%; Score 15.8; DB 6; Length 461;
Best Local Similarity 89.5%; Pred. No. 7.1e+03;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GTCGGTGCAGGACGTGACA 19

Db 9 GTCGCTGCTGGACGTGACA 27

RESULT 12
LOCUS PPU80658 698 bp DNA linear PLN 05-JAN-1999
DEFINITION Peniophora pini 25S nuclear ribosomal RNA gene, partial sequence.
ACCESSION U80658
VERSION U80658.1 GI:4098668
KEYWORDS
SOURCE Peniophora pini.
ORGANISM Peniophora pini
Eukaryota; Fungi; Basidiomycota; Hymenomycetes; Homobasidiomycetes;
Aphyllophorales; Lachnociadiaceae; Peniophora.

REFERENCE 1 (bases 1 to 698)
Hallenberg, N. and Parmasto, E.
AUTHORS Hallenberg, N. and Parmasto, E.
TITLE Phylogenetic studies in species of Corticiaceae growing on branches
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 698)
Hallenberg, N.
AUTHORS Hallenberg, N.
TITLE Direct Submission
JOURNAL Submitted (02-DEC-1996) Systematic Botany, Botanical Institute,
Carl Skottsbergs gata 22, Gothenburg S-413 19, Sweden

FEATURES Location/Qualifiers
source 1..698
/organism="Peniophora pini"
/strain="FCUG 2399"
/db_xref="taxon:55353"
<1..>698
/product="25S ribosomal RNA"
BASE COUNT 109 a 144 c 102 g 124 t 219 others
ORIGIN

Query Match 83.2%; Score 15.8; DB 8; Length 698;
Best Local Similarity 89.5%; Pred. No. 6.9e+03;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GTCGGTGCAGGACGTGACA 19

Db 56 GTCGCTGCTGGACGTGACA 74

RESULT 13
LOCUS CRU80647 699 bp DNA linear PLN 05-JAN-1999
DEFINITION Corticium roseum 25S nuclear ribosomal RNA gene, partial sequence.
ACCESSION U80647
VERSION U80647.1 GI:4098657
KEYWORDS
SOURCE Corticium roseum.
ORGANISM Corticium roseum
Eukaryota; Fungi; Basidiomycota; Hymenomycetes; Homobasidiomycetes;
Aphyllophorales; Corticiaceae; Corticium.

REFERENCE 1 (bases 1 to 699)
Hallenberg, N. and Parmasto, E.
AUTHORS Hallenberg, N. and Parmasto, E.
TITLE Phylogenetic studies in species of Corticiaceae growing on branches
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 699)

AUTHORS Hallenberg, N.
TITLE Direct Submission
JOURNAL Submitted (02-DEC-1996) Systematic Botany, Botanical Institute,
Carl Skottsbergs gata 22, Gothenburg S-413 19, Sweden
FEATURES Location/Qualifiers
source 1..699
/organism="Corticium roseum"
/strain="FCUG 2598"
/db_xref="taxon:55342"

BASE COUNT 113 a 140 c 97 g 130 t 219 others
ORIGIN

Query Match 83.2%; Score 15.8; DB 8; Length 699;
Best Local Similarity 89.5%; Pred. No. 6.9e+03;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GTCGGTGCAGGACGTGACA 19

Db 56 GTCGCTGCTGGACGTGACA 74

RESULT 14
LOCUS PPU80657 807 bp DNA linear PLN 05-JAN-1999
DEFINITION Peniophora piceae 25S nuclear ribosomal RNA gene, partial sequence.
ACCESSION U80657
VERSION U80657.1 GI:4098667
KEYWORDS
SOURCE Peniophora piceae.
ORGANISM Peniophora piceae
Eukaryota; Fungi; Basidiomycota; Hymenomycetes; Homobasidiomycetes;
Aphyllophorales; Lachnociadiaceae; Peniophora.

REFERENCE 1 (bases 1 to 807)
Hallenberg, N. and Parmasto, E.
AUTHORS Hallenberg, N. and Parmasto, E.
TITLE Phylogenetic studies in species of Corticiaceae growing on branches
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 807)
Hallenberg, N.
AUTHORS Hallenberg, N.
TITLE Direct Submission
JOURNAL Submitted (02-DEC-1996) Systematic Botany, Botanical Institute,
Carl Skottsbergs gata 22, Gothenburg S-413 19, Sweden

FEATURES Location/Qualifiers
source 1..807
/organism="Peniophora piceae"
/strain="FCUG 2306"
/db_xref="taxon:55352"
<1..>807
/product="25S ribosomal RNA"
BASE COUNT 133 a 181 c 119 g 155 t 219 others
ORIGIN

Query Match 83.2%; Score 15.8; DB 8; Length 807;
Best Local Similarity 89.5%; Pred. No. 6.8e+03;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GTCGGTGCAGGACGTGACA 19

Db 165 GTCGCTGCTGGACGTGACA 183

RESULT 15
LOCUS PPU80660 812 bp DNA linear PLN 05-JAN-1999
DEFINITION Peniophora proxima 25S nuclear ribosomal RNA gene, partial sequence.
ACCESSION U80660
VERSION U80660.1 GI:4098670
KEYWORDS
SOURCE Peniophora proxima.
ORGANISM Peniophora proxima
Eukaryota; Fungi; Basidiomycota; Hymenomycetes; Homobasidiomycetes;

REFERENCE 1 Aphyllophorales; Lachnocladiaceae; Peniophora.
 1 (bases 1 to 812)
 AUTHORS Hallenberg, N. and Barnasto, E.
 TITLE Phylogenetic studies in species of Corticiaceae growing on branches
 JOURNAL Unpublished
 REFERENCE 2 (bases 1 to 812)

AUTHORS Hallenberg, N.
 TITLE Direct Submission
 JOURNAL Submitted (02-DEC-1996) Systematic Botany, Botanical Institute,
 Carl Skottsbergs gata 22, Gothenburg S-413 19, Sweden

FEATURES Location/Qualifiers

Source
 1..812
 /organism="Peniophora proxima"
 /strain="FCUG 1795"
 /db_xref="taxon:55355"
 <1..>812

BASE COUNT 135 a 181 c 119 g 158 t 219 others
 ORIGIN

Query Match 83.2%; Score 15.8; DB 8; Length 812;
 Best Local Similarity 89.5%; Pred. No. 6.8e+03;
 Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 GTCGGTGCAGGACGTGACA 19
 |||||
 Db 171 GTCGGTGCAGGCGCGACA 189

Search completed: January 3, 2003, 23:54:35
 Job time : 265.12 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 3, 2003, 22:52:02 ; Search time 31.7666 Seconds
(without alignments)
1346.950 Million cell updates/sec

Title: us-09-787-562-2

Perfect score: 19
Sequence: 1 gtcggtgcaggacgtgaca 19

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 2185239 seqs, 1125999159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0

Maximum DB seq length: 20000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

N_Geneseq_101002:*

- 1: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1980.DAT.*
- 2: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1981.DAT.*
- 3: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1982.DAT.*
- 4: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1983.DAT.*
- 5: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1984.DAT.*
- 6: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1985.DAT.*
- 7: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1986.DAT.*
- 8: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1987.DAT.*
- 9: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1988.DAT.*
- 10: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1989.DAT.*
- 11: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1990.DAT.*
- 12: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1991.DAT.*
- 13: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1992.DAT.*
- 14: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1993.DAT.*
- 15: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1994.DAT.*
- 16: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1995.DAT.*
- 17: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1996.DAT.*
- 18: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1997.DAT.*
- 19: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1998.DAT.*
- 20: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1999.DAT.*
- 21: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2000.DAT.*
- 22: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT.*
- 23: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT.*
- 24: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	19	100.0	19	21	AA11994
2	19	100.0	25	21	AA11993
3	16	84.2	517	23	AA80329
4	16	84.2	633	23	AA80330
5	16	84.2	1624	21	AA80386
6	15.8	83.2	461	24	ABK80953
7	15.8	83.2	1077	21	AAA30607
8	15.8	83.2	1077	21	AAA30720
9	15.8	83.2	1545	22	AAF61096

c	10	15.8	83.2	1815	20	AAZ06792	BCR673 (7 transmem
c	11	15.4	81.1	240	21	ABQ62698	Myobacterium tube
c	12	15.4	81.1	1202	23	AA86562	DNA encoding novel
c	13	15.4	81.1	1332	22	AAF60970	P. putida KT2440-a
c	14	15.4	81.1	4403765	22	AAI99683	Myobacterium tube
c	15	15.4	81.1	4411529	22	AAI99682	Myobacterium tube
c	16	15	78.9	18	16	AAQ99459	Hypoxia-inducible
c	17	15	78.9	18	21	AA12053	Murine PGK1 derive
c	18	15	78.9	18	21	AA12054	Murine promoter OB
c	19	15	78.9	19	22	AAF85326	Nucleotide fragmen
c	20	15	78.9	21	22	AAH42134	HRE element from t
c	21	15	78.9	24	16	AAQ99458	Hypoxia-inducible
c	22	15	78.9	24	20	AAZ11422	Murine HRE mPK DN
c	23	15	78.9	24	21	AA12007	Murine hypoxic res
c	24	15	78.9	24	22	AAH88980	Synapsin gene SIL
c	25	15	78.9	43	22	AAH42138	HRE-containing enh
c	26	15	78.9	72	20	AAZ11440	Murine PGK HRE Hf
c	27	15	78.9	72	21	AA12023	Synapsin gene SIL
c	28	15	78.9	86	22	AAH42139	ETAV U3 enhancer r
c	29	15	78.9	86	22	AAH42139	ETAV U3 enhancer r
c	30	15	78.9	100	21	AA12060	Synapsin gene SIL
c	31	15	78.9	114	21	AA12061	Synapsin gene SIL
c	32	15	78.9	123	22	AAH42140	PGK derived enhanc
c	33	15	78.9	129	22	AAH42142	PGK derived enhanc
c	34	15	78.9	225	20	AAZ11399	Murine PGK HRE der
c	35	15	78.9	225	21	AA11997	Promoter OBHrel us
c	36	15	78.9	229	20	AAZ11398	Murine PGK HRE der
c	37	15	78.9	229	21	AA11996	Murine PGK HRE der
c	38	15	78.9	237	21	AA12001	HIV derived synthe
c	39	15	78.9	242	20	AAZ07789	Murine PGK HRE der
c	40	15	78.9	242	21	AA12016	Promoter OBHrel us
c	41	15	78.9	243	20	AAZ11397	Murine PGK HRE der
c	42	15	78.9	243	21	AA11995	Murine PGK fragmen
c	43	15	78.9	245	20	AAZ11438	HIV derived synthe
c	44	15	78.9	249	21	AA11999	Murine PGK HRE der
c	45	15	78.9	269	20	AAZ11439	Synthetic promoter

ALIGNMENTS

RESULT 1
AA11994
ID AAA11994 standard; DNA; 19 BP.

AC AA11994;

XX
XX
DT 14-AUG-2000 (first entry)

DE Murine PGK HRE truncated P18 DNA sequence.

DE HRE; hypoxia response element; hypoxia-inducible factor; HIF; vasotropic;
KW cardiant; cyclostatic; antiarthritis; gene therapy; ischaemia; arthritis;
KW cardiovascular disease; peripheral arterial disease; cancer;
KW phosphoglycerate kinase; PGK; murine; ds.
OS Mus sp.

PN WO200017371-A1.

PD 30-MAR-2000.

XX 22-SEP-1999; 99WO-GB03181.

XX 23-SEP-1998; 98WO-CB02885.

XX 28-JAN-1999; 99GB-0001906.

XX 16-FEB-1999; 99GB-0003538.

PA (OXFO-) OXFORD BIOMEDICA UK LTD.

XX Binley KM, Naylor S;

XX WPI; 2000-283595/24.

Mon Jan 6 15:20:22 2003

XX Novel polynucleotide constructs comprising at least two repeats of a
PT hypoxia response element useful for driving expression of nucleic acids
PT of interest in a cell under hypoxic conditions -
XX Disclosure; Page 11; 155pp; English.
XX
XX This invention describes novel polynucleotide comprising at least 2
CC repeats of a hypoxia response element (HRE), where the hypoxia-inducible
CC factor (HIF) consensus binding sites within each of the 2 repeats are
CC separated by a spacer of at least 20 contiguous nucleotides. The products
CC of the invention have vasotropic, cardiant, cytostatic and antiarthritic
CC activity and can be used for gene therapy. The polynucleotides are useful
CC for delivering nucleic acids of interest to mammalian cells. Lentiviral
CC vectors are responsive to hypoxic agents and to agents that mimic
CC hypoxia. This regulation can be harnessed in vitro to enhance the
CC production of the vector and can be used in vivo to regulate gene
CC expression in response to a physiological signal. The vectors have
CC utility in disease, where ischaemia, including hypoxia, is a feature,
CC e.g. cardiovascular disease, peripheral arterial disease, cancer and
CC arthritis. The novel regulatory construct is capable of driving very high
CC levels of transcription under conditions of hypoxia whilst providing only
CC low basal levels of transcription under normal oxygen conditions. The
CC polynucleotide construct targets cells within a tumor mass that are under
CC conditions of hypoxia without affecting normal surrounding tissue. This
CC sequence represents a murine phosphoglycerate kinase (PGK) HRE truncated
CC p18 DNA fragment as described in the method of the invention.
XX
XX Sequence 19 BP; 4 A; 4 C; 8 G; 3 T; 0 other;
SQ
Query Match 100.0%; Score 19; DB 21; Length 19;
Best Local Similarity 100.0%; Pred. No. 6.4;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GTCGGTGCAGGACGTGACA 19
DB 1 GTCGGTGCAGGACGTGACA 19
|||||
RESULT 2
AAAL1993
ID AAAL1993 standard; DNA; 25 BP.
XX
XX AAAL1993;
XX
XX 14-AUG-2000 (first entry)
XX
XX Murine PGK HRE p42 DNA sequence.
XX
XX HRE; hypoxia response element; hypoxia-inducible factor; HIF; vasotropic;
KW cardiant; cytostatic; antiarthritic; gene therapy; ischaemia; arthritis;
KW cardiovascular disease; peripheral arterial disease; cancer;
KW phosphoglycerate kinase; PGK; murine; ds.
XX
XX Mus sp.
XX
XX WO200017371-A1.
XX
XX 30-MAR-2000.
XX
XX 22-SEP-1999; 99WO-GB03181.
XX
XX 23-SEP-1998; 98WO-GB02885.
XX
XX 28-JAN-1999; 99GB-0001906.
XX
XX 16-FEB-1999; 99GB-0003538.
XX
XX (OXFO-) OXFORD BIOMEDICA UK LTD.
XX
XX Binley KM, Naylor S;
XX
XX WPI; 2000-283595/24.
XX
XX Novel polynucleotide constructs comprising at least two repeats of a

PT hypoxia response element useful for driving expression of nucleic acids
PT of interest in a cell under hypoxic conditions -
XX Disclosure; Page 11; 155pp; English.
XX
XX This invention describes novel polynucleotide comprising at least 2
CC repeats of a hypoxia response element (HRE), where the hypoxia-inducible
CC factor (HIF) consensus binding sites within each of the 2 repeats are
CC separated by a spacer of at least 20 contiguous nucleotides. The products
CC of the invention have vasotropic, cardiant, cytostatic and antiarthritic
CC activity and can be used for gene therapy. The polynucleotides are useful
CC for delivering nucleic acids of interest to mammalian cells. Lentiviral
CC vectors are responsive to hypoxic agents and to agents that mimic
CC hypoxia. This regulation can be harnessed in vitro to enhance the
CC production of the vector and can be used in vivo to regulate gene
CC expression in response to a physiological signal. The vectors have
CC utility in disease, where ischaemia, including hypoxia, is a feature,
CC e.g. cardiovascular disease, peripheral arterial disease, cancer and
CC arthritis. The novel regulatory construct is capable of driving very high
CC levels of transcription under conditions of hypoxia whilst providing only
CC low basal levels of transcription under normal oxygen conditions. The
CC polynucleotide construct targets cells within a tumor mass that are under
CC conditions of hypoxia without affecting normal surrounding tissue. This
CC sequence represents a murine phosphoglycerate kinase (PGK) HRE p24 DNA
CC fragment as described in the method of the invention.
XX
XX Sequence 25 BP; 6 A; 6 C; 9 G; 4 T; 0 other;
SQ
Query Match 100.0%; Score 19; DB 21; Length 25;
Best Local Similarity 100.0%; Pred. No. 6.5;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GTCGGTGCAGGACGTGACA 19
DB 4 GTCGGTGCAGGACGTGACA 22
|||||
RESULT 3
AAS80329
ID AAS80329 standard; cDNA; 517 BP.
XX
XX AAS80329;
XX
XX 13-FEB-2002 (first entry)
XX
XX DNA encoding novel human diagnostic protein #16133.
XX
XX Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX
XX Homo sapiens.
XX
XX WO200175067-A2.
XX
XX 11-OCT-2001.
XX
XX 30-MAR-2001; 2001WO-US08631.
XX
XX 31-MAR-2000; 2000US-0540217.
XX
XX 23-AUG-2000; 2000US-0649167.
XX
XX (HYSE-) HYSEQ INC.
XX
XX Drmanac RT, Liu C, Tang YT;
XX
XX WPI; 2001-639362/73.
XX
XX P-PSDB; ABG16142.
XX
XX New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity -
XX

PS Claim 1; SEQ ID No 16133; 103pp; English.

XX The invention relates to isolated polynucleotide (I) and polypeptide (II) sequences. (I) is useful as hybridisation probes, polymerase chain reaction (PCR) primers, oligomers, and for chromosome and gene mapping, and in recombinant production of (II). The polynucleotides are also used in diagnostics as expressed sequence tags for identifying expressed genes. (I) is useful in gene therapy techniques to restore normal activity of (II) or to treat disease states involving (II). (II) is useful for generating antibodies against it, detecting or quantitating a polypeptide in tissue, as molecular weight markers and as a food supplement. (II) and its binding partners are useful in medical imaging of sites expressing (II). (I) and (II) are useful for treating disorders involving aberrant protein expression or biological activity. The polypeptide and polynucleotide sequences have applications in diagnostics, forensics, gene mapping, identification of mutations responsible for genetic disorders or other traits to assess biodiversity and to produce other types of data and products dependent on DNA and amino acid sequences. AAS64197-AAS94564 represent novel human diagnostic coding sequences of the invention.

CC Note: The sequence data for this patent did not appear in the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published_pct_sequences.

XX SQ Sequence 517 BP; 101 A; 137 C; 174 G; 105 T; 0 other;

Query Match 84.2%; Score 16; DB 23; Length 517;
Best Local Similarity 100.0%; Pred. No. 2.1e+02;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GTCGGTGCAGGACGTG 16
|||||
Db 412 GTCGGTGCAGGACGTG 427

RESULT 4

AAS80330 ID AAS80330 standard; cDNA; 633 BP.

XX AAC80330; AC

XX 13-FEB-2002 (first entry)

DE DNA encoding novel human diagnostic protein #16134.

XX Human; chromosome mapping; gene mapping; gene therapy; forensic; food supplement; medical imaging; diagnostic; genetic disorder; ss.

OS Homo sapiens.

XX WO200175067-A2.

PN 11-OCT-2001.

XX 30-MAR-2001; 2001WO-US08631.

XX 31-MAR-2000; 2000US-0540217.

XX 23-AUG-2000; 2000US-0649167.

XX (HYSE-) HYSEQ INC.

XX Drmanac RT, Liu C, Tang YT;

XX WPI; 2001-639362/73.

XX P-PSDB; ABG16143.

XX New isolated polynucleotide and encoded polypeptides, useful in diagnostics, forensics, gene mapping, identification of mutations responsible for genetic disorders or other traits and to assess biodiversity.

XX Claim 1; SEQ ID No 16134; 103pp; English.

CC The invention relates to isolated polynucleotide (I) and polypeptide (II) sequences. (I) is useful as hybridisation probes, polymerase chain reaction (PCR) primers, oligomers, and for chromosome and gene mapping, and in recombinant production of (II). The polynucleotides are also used in diagnostics as expressed sequence tags for identifying expressed genes. (I) is useful in gene therapy techniques to restore normal activity of (II) or to treat disease states involving (II). (II) is useful for generating antibodies against it, detecting or quantitating a polypeptide in tissue, as molecular weight markers and as a food supplement. (II) and its binding partners are useful in medical imaging of sites expressing (II). (I) and (II) are useful for treating disorders involving aberrant protein expression or biological activity. The polypeptide and polynucleotide sequences have applications in diagnostics, forensics, gene mapping, identification of mutations responsible for genetic disorders or other traits to assess biodiversity and to produce other types of data and products dependent on DNA and amino acid sequences. AAS64197-AAS94564 represent novel human diagnostic coding sequences of the invention.

CC Note: The sequence data for this patent did not appear in the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published_pct_sequences.

XX SQ Sequence 633 BP; 147 A; 166 C; 189 G; 131 T; 0 other;

Query Match 84.2%; Score 16; DB 23; Length 633;
Best Local Similarity 100.0%; Pred. No. 2.1e+02;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GTCGGTGCAGGACGTG 16
|||||
Db 447 GTCGGTGCAGGACGTG 462

RESULT 5

AAC79886 ID AAC79886 standard; cDNA; 1624 BP.

XX AAC79886; AC

XX 09-FEB-2001 (first entry)

DE Human secreted protein encoding cDNA for gene 38.

XX Human; secreted protein; cytostatic; antiarthritic; antiasthmatic; immunosuppressive; antiarteriosclerotic; antiinflammatory; neurotropic; neuroprotective; antidiabetic; tranquiliser; vulnerary; antibacterial; antipsoriatic; antiarrhythmic; antirheumatic; cardiant; anti-HIV; autoimmune disorder; allergic condition; cardiovascular disorder; cancer; neurological disease; tissue repair; ss.

XX Homo sapiens.

XX WO2000055176-A2.

XX 21-SEP-2000.

XX 09-MAR-2000; 2000WO-US06057.

XX 12-MAR-1999; 99US-0124142.

XX 11-JUN-1999; 99US-0138597.

XX 03-DEC-1999; 99US-0168666.

XX (HUMA-) HUMAN GENOME SCI INC.

XX Rosen CA, Ruben SM, Komatsoulis G;

XX WPI; 2000-638176/61.

XX P-PSDB; AAB44867.

XX Novel 49 human secreted proteins useful for diagnosis, prevention and treatment of disorders including neurological, cell proliferative, cardiovascular, and autoimmune/inflammatory disorders and microbial infections.

XX Claim 1a; Page 355; 405pp; English.

XX This invention describes a novel isolated polypeptide (I) comprising an

XX amino acid sequence at least 95 % identical to a polypeptide sequence

XX selected from 49 polypeptides encoded by polynucleotide sequences

XX included in American Type Culture Collection (ATCC) deposit number

XX 203917, defined in the specification. The products of the invention have

XX cytostatic, antiarthritic, antiasthmatic, immunosuppressive, neurotropic;

XX antiarteriosclerotic, antiinflammatory, neuroprotective, antidiabetic,

XX tranquiliser, vulnerary, antibacterial, antiporiatic, antiarrhythmic,

XX antirheumatic, cardiant and anti-HIV activity. (I) or a nucleic acid (II)

XX encoding (I) is useful for preventing, treating or ameliorating a medical

XX condition and for diagnosing a pathological condition or susceptibility

XX to the condition. (I) is useful for identifying a binding partner which

XX affects the activity of the polypeptide and for identifying an activity

XX in a biological sample. (I), (II) or an antibody (IV) specific to (I) is

XX also useful for treating or preventing a disease, disorder or condition

XX associated with aberrant expression of (I). Diseases treated or diagnosed

XX include immune disorders such as autoimmune diseases, blood protein

XX disorders, anemia, allergic reactions and conditions such as asthma,

XX organ rejection or graft-versus-host disease, inflammation, hyper

XX proliferative disorders, cardiovascular disorders such as arterioarterial

XX fistula, arrhythmias, arteriosclerosis, coronary thrombosis, organ

XX regeneration, cancer, neovascular glaucoma, diabetic retinopathy,

XX rheumatoid arthritis, psoriasis, diseases associated with increased

XX apoptosis that include acquired immunodeficiency syndrome (AIDS),

XX neurological diseases such as Parkinson's disease, viral, bacterial,

XX fungal or parasitic diseases. They are also used to repair, replace or

XX protect tissue damage by congenital defects, to treat trauma, in surgery,

XX including cosmetic plastic surgery, to treat fibrosis, reperfusion injury

XX or systemic cytokine damage, to stimulate chondrocyte growth, to prevent

XX skin aging due to sunburn, to change a mammal's mental state or physical

XX state by influencing biorhythms, cardiac rhythms, depression, memory,

XX stress and for accelerating wound healing. (I), (II) and/or their agonist

XX or antagonist are useful as food additives or preservatives to increase

XX or decrease storage capabilities, fat content, lipid, protein,

XX carbohydrate, vitamin, mineral or other nutritional components. (I) is

XX useful for screening therapeutic compounds. (II) is useful in forensic

XX biology for detecting DNA sequences and as diagnostic probes for

XX detecting the presence of specific mRNA in a particular cell type.

XX Sequence 1624 BP; 303 A; 506 C; 503 G; 312 T; 0 other;

Query Match 84.2%; Score 16; DB 21; Length 1624;

Best Local Similarity 100.0%; Pred. No. 2.1e+02;

Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GTCGGTGCAGGACGTG 16

DB 666 GTCGGTGCAGGACGTG 681

RESULT 6

ABK80953

ID ABK80953 standard; DNA; 461 BP.

XX AC ABK80953;

XX DT 13-AUG-2002 (first entry)

XX DE Bacillus clausii genomic sequence tag (GST) #3796.

XX KW Differential gene expression; genomic sequenced tag; GST;

XX KW altered culture condition; environmental stress;

XX KW physiological provocation; ds.

XX OS Bacillus clausii.

XX PN WO200229113-A2.

XX XX 11-APR-2002.

XX PD

PF 05-OCT-2001; 2001WO-US31437.

XX PR 06-OCT-2000; 2000US-0680598.

PR 27-MAR-2001; 2001US-279526P.

XX PA (NOVO) NOVOZYMES BIOTECH INC.

PA (NOVO) NOVOZYMES AS.

XX PI Berka R, Clausen IG;

XX WPI; 2002-416684/44.

XX PT Monitoring differential expression of several genes in first Bacillus

PT cell relative to expression of same genes in one or more second

PT Bacillus cells, by using substrate containing Bacillus genomic

PT sequenced tag array

XX PS Claim 11; SEQ ID NO 8244; 200pp; English.

XX The invention describes a method of monitoring differential expression of

XX genes in a first Bacillus cell relative to expression of the genes in

XX other Bacillus cells, comprising hybridising labelled nucleic acid probes

XX isolated from Bacillus cells to a substrate containing array of Bacillus

XX genomic sequenced tags (GST), examining the array, and determining

XX relative gene expression by an observed hybridisation reporter signal of

XX a spot in the array. The method is useful for measuring the expression of

XX genes in a first Bacillus cell relative to expression of the same genes

XX in one or more second Bacillus cells. The method is useful for monitoring

XX global expression of several genes from a Bacillus cell, discovering new

XX in which Bacillus cells adapt to changes in culture conditions,ive

XX environmental stress or other physiological provocation. Extensive

XX follow-up characterisation is unnecessary, when one spot on an array

XX equals one gene or one open reading frame, since sequence information is

XX available. This sequence represents a genomic sequence tag (GST) used in

XX the method of the invention.

XX Note: The sequence data for this patent did not form part of the printed

XX specification, but was obtained in electronic format directly from WIPO

XX at

XX ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 461 BP; 124 A; 94 C; 118 G; 125 T; 0 other;

Query Match 83.2%; Score 15.8; DB 24; Length 461;

Best Local Similarity 89.5%; Pred. No. 2.6e+02;

Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GTCGGTGCAGGACGTGACA 19

DB 9 GTCGGTGCAGGACGTGACA 27

RESULT 7

AAA30607

ID AAA30607 standard; cDNA; 1077 BP.

XX AC AAA30607;

XX DT 21-AUG-2000 (first entry)

XX DE Human G protein-coupled receptor GPR20 cDNA.

XX KW G protein-coupled receptor; GPCR; constitutively active;

XX KW intracellular loop 3; transmembrane domain 6; drug screening;

XX KW agonist; antagonist; ss.

XX OS Homo sapiens.

XX PN WO200022129-A1.

XX XX 20-APR-2000.

XX PD

```

XX PF 12-OCT-1999; 99WO-US23938.
XX PR 13-OCT-1998; 98US-0170496.
XX PA (AREN-) ARENA PHARM INC.
XX PI Behan DP, Chalmers DT, Liaw CW;
XX DR WPI; 2000-329165/28.
XX DR P-PSDB; AAY90620.
XX PT Non-endogenous constitutively activated human G protein-coupled
XX PT receptors, useful for identifying agonists for use as pharmaceutical
XX PT agents
XX PS Example 1; Page 131; 341pp; English.
XX CC The invention relates to constitutively active, non-endogenous versions
XX CC of endogenous human orphan G protein-coupled receptors (GPCRs, AAY90643-
XX CC AAY90677 and AAY90683-Y90687), and to DNA encoding them (AAA30709-A30743
XX CC and AAA30775-A30779). The mutant proteins of the invention contain a
XX CC mutation in a portion of the protein comprising intracellular loop 3
XX CC (IC3) and transmembrane domain 6 (TM6). A non-endogenous amino acid, X,
XX CC is substituted for an endogenous residue in TM6 to form a sequence
XX CC acids N-terminal of an endogenous residue in IC3 at a position 16 amino
XX CC X-(AA)15-Pro. The endogenous proline in TM6 is selected from Lys, His, Arg
XX CC or Ala, and is preferably Lys. When the endogenous residue at this
XX CC position is Lys, this residue is replaced by His, Arg or preferably Ala.
XX CC The 15 amino acid stretch between the substituted amino acid and the Pro
XX CC may be endogenous, non-endogenous, or a mixture of endogenous and
XX CC non-endogenous residues. The constitutively active GPCRs are useful for
XX CC identifying antagonists, agonists and partial agonists for use as
XX CC pharmaceutical agents. The mutant proteins are also useful in research
XX CC settings for elucidating the roles of the receptors in normal and
XX CC diseased conditions. Antagonists for a particular GPCR are useful for
XX CC treating diseases and disorders associated with that receptor. Because
XX CC the novel mutant GPCRs are constitutively active, they can be used
XX CC directly for screening of compounds without the need for endogenous
XX CC ligands. The present sequence represents cDNA encoding a human wild-type
XX CC GPCR used in an exemplification of the invention. This was cloned and
XX CC subjected to site-directed mutagenesis (SDM) to generate DNA encoding
XX CC the corresponding mutant of the invention.
XX SQ Sequence 1077 BP; 142 A; 399 C; 320 G; 216 T; 0 other;

Query Match 83.2%; Score 15.8; DB 21; Length 1077;
Best Local Similarity 89.5%; Pred. No. 2.6e+02;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GTCGGTGCAGGACGTGACA 19
Db 549 GTCGGTGTGGGCGTGACA 567

RESULT 8
AAA30720
ID AAA30720 standard; DNA; 1077 BP.
XX AC
XX AC AAA30720;
XX DT 21-AUG-2000 (first entry)
XX DE DNA encoding human mutant G protein-coupled receptor GPR20 (M240K).
XX KW G protein-coupled receptor; GPCR; constitutively active;
XX KW intracellular loop 3; transmembrane domain 6; drug screening;
XX KW agonist; antagonist; mutant; ss.
XX OS Homo sapiens.
XX OS Synthetic.
XX PN WO200022129-A1.

```

```

XX PD 20-APR-2000.
XX PF 12-OCT-1999; 99WO-US23938.
XX PR 13-OCT-1998; 98US-0170496.
XX PA (AREN-) ARENA PHARM INC.
XX PI Behan DP, Chalmers DT, Liaw CW;
XX DR WPI; 2000-329165/28.
XX DR P-PSDB; AAY90654.
XX PT Non-endogenous constitutively activated human G protein-coupled
XX PT receptors, useful for identifying agonists for use as pharmaceutical
XX PT agents
XX PS Example 2; Page 237; 341pp; English.
XX CC The invention relates to constitutively active, non-endogenous versions
XX CC of endogenous human orphan G protein-coupled receptors (GPCRs, AAY90643-
XX CC AAY90677 and AAY90683-Y90687), and to DNA encoding them (AAA30709-A30743
XX CC and AAA30775-A30779). The mutant proteins of the invention contain a
XX CC mutation in a portion of the protein comprising intracellular loop 3
XX CC (IC3) and transmembrane domain 6 (TM6). A non-endogenous amino acid, X,
XX CC is substituted for an endogenous residue in TM6 to form a sequence
XX CC acids N-terminal of an endogenous residue in IC3 at a position 16 amino
XX CC X-(AA)15-Pro. The endogenous proline in TM6 is selected from Lys, His, Arg
XX CC or Ala, and is preferably Lys. When the endogenous residue at this
XX CC position is Lys, this residue is replaced by His, Arg or preferably Ala.
XX CC The 15 amino acid stretch between the substituted amino acid and the Pro
XX CC may be endogenous, non-endogenous, or a mixture of endogenous and
XX CC non-endogenous residues. The constitutively active GPCRs are useful for
XX CC identifying antagonists, agonists and partial agonists for use as
XX CC pharmaceutical agents. The mutant proteins are also useful in research
XX CC settings for elucidating the roles of the receptors in normal and
XX CC diseased conditions. Antagonists for a particular GPCR are useful for
XX CC treating diseases and disorders associated with that receptor. Because
XX CC the novel mutant GPCRs are constitutively active, they can be used
XX CC directly for screening of compounds without the need for endogenous
XX CC ligands. Sequences AAA30709- AAA30743 and AAA30775-A30779 represent DNAs
XX CC encoding the mutant human GPCRs of the invention.
XX SQ Sequence 1077 BP; 143 A; 399 C; 320 G; 215 T; 0 other;

Query Match 83.2%; Score 15.8; DB 21; Length 1077;
Best Local Similarity 89.5%; Pred. No. 2.6e+02;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GTCGGTGCAGGACGTGACA 19
Db 549 GTCGGTGTGGGCGTGACA 567

RESULT 9
AAF61096
ID AAF61096 standard; DNA; 1545 BP.
XX AC
XX AC AAF61096;
XX DT 16-MAY-2001 (first entry)
XX DE P. putida KT2440-associated DNA ORF11200.
XX KW Transgenic plant; detection; probe; amplification; vaccine carrier;
XX KW microbial production strain; biological remediation; ds.
XX OS Pseudomonas putida.
XX PN DE19935088-A1.
XX PN 01-FEB-2001.

```

Mon Jan 6 15:20:22 2003

```

XX PF 27-JUL-1999; 99DE-1035088.
XX PR 27-JUL-1999; 99DE-1035088.
XX (TIGR-) TIGR INST GENOMIC RES.
XX (QIITA-) QIITA GEN.
XX (GBF-) GBS BIOTECHNOLOGISCHE FORSCHUNG MBH.
XX (DKFZ-) DKFZ DEUT KREBSFORSCHUNGSZENTRUM.
XX (MEDI-) MEDIZINISCHE HOCHSCHULE HANNOVER.
XX WPI; 2001-192469/20.
XX
XX PT New DNA sequences specific for Pseudomonas putida KT2440, useful as
XX safe genetic engineering host, allow detection in presence of other
XX related bacteria.
XX
XX PS Claim 1a; Page 149; 158pp; German.
XX
XX This invention describes novel DNA sequences (I) for specific detection
XX of Pseudomonas putida KT2440. The invention also describes (1)
XX recombinant expression vector containing (I); (2) prokaryotic or
XX eukaryotic cells transformed or transfected with (I) or the vector of
XX (1); (3) production of expression products by culturing cells of (2);
XX (4) expression products, or their fragments, of (1) and synthetic
XX proteins or peptides with the same sequences (A); (5) poly- or
XX mono-clonal antibodies (Ab) that react specifically with (A); (6)
XX hybridoma cells that produce the monoclonal Ab of (5); (7) transgenic
XX plants that contain transformed or transfected cells of (2); (8)
XX detecting KT2440 using a labeled (I) or Ab as probe; and (9) DNA chips
XX carrying one or more (I), and their fragments, are used as probes
XX to detect and isolate full-length cDNAs and/or to amplify such cDNAs by
XX polymerase chain reaction, and for production of transgenic plants. (I),
XX or antibodies that recognize their expression products, are used for
XX detecting the presence of KT2440, particularly in presence of other,
XX even closely related, bacteria. KT2440 is one of the bacteria classified
XX as safe, by the National Institutes of Health, for genetic engineering
XX work, e.g. as microbial production strains, for biological remediation
XX and as vaccine carriers. (I) are exclusive to KT2440 with no significant
XX homology with sequences in other bacteria (specifically the closely
XX related pathogen P. aeruginosa). Compared with other 'safe' bacteria, it
XX has greater catabolic activity and better survival in, and adaptation to,
XX the rhizosphere and soil.
XX
XX SQ Sequence 1545 BP; 255 A; 491 C; 496 G; 303 T; 0 other;

```

```

FT FT
FT FT
XX XX
XX PN WO9946296-A1.
XX
XX PD 16-SEP-1999.
XX
XX PF 01-MAR-1999; 99WO-US04422.
XX
XX PR 11-MAR-1998; 98US-0041545.
XX
XX PA (SMK ) SMITHKLINE BEECHAM CORP.
XX
XX PI Halsey WS, Mao JY, Sathe GM;
XX
XX DR WPI; 1999-551366/46.
XX
XX PS P-PSDB; AAY39292.
XX
XX PT New 7-transmembrane G-protein coupled receptor designated ECR673 used
XX in therapy and for treating diseases.
XX
XX PS Claim 2; Page 30; 43pp; English.
XX
XX This sequence is the gene of a human 7-transmembrane G-protein coupled
XX receptor designated ECR673. Many medically significant biological
XX processes are mediated by G-protein coupled receptors involved in signal
XX transduction pathways. ECR673 polypeptides may be used for identifying
XX agonists and antagonists/inhibitors and for detecting diseases
XX associated with inappropriate ECR673 activity or levels. ECR673
XX polypeptides and polynucleotides, agonists, antagonists and antibodies
XX are used to treat diseases, including infections such as bacterial,
XX fungal, protozoan and viral infections, especially HIV-1 or HIV-2; pain;
XX cancer; diabetes; obesity; anorexia; bulimia; asthma;
XX Parkinson's disease; acute heart failure; hypertension; hypotension;
XX urinary retention; osteoporosis; angina pectoris; myocardial infarction;
XX stroke; ulcers; allergies; benign prostatic hypertrophy; migraine;
XX vomiting; psychotic and neurological disorders, including anxiety,
XX schizophrenia, manic depression, depression, delirium, dementia, and
XX severe mental retardation; and dyskinesias such as Huntington's disease
XX or Gilles de la Tourette's syndrome. The polynucleotide is also useful as a
XX source of primers and probes, and also for detecting the above diseases.
XX
XX SQ Sequence 1815 BP; 274 A; 623 C; 555 G; 362 T; 1 other;

```

```

Query Match 83.2%; Score 15.8; DB 20; Length 1815;
Best Local Similarity 89.5%; Pred. NO. 2.7e+02;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```

```

QY 1 GTCGGTGCAGGACGTGACA 19
|||||
DB 1169 GTCGGTGCAGGACGTGACA 1187

```

```

RESULT 11
ABQ62698/c
ID ABQ62698 standard; DNA; 240 BP.
XX
XX AC ABQ62698;
XX
XX DT 16-AUG-2002 (first entry)
XX
XX DE Mycobacterium tuberculosis BAC vector clone Rv17.
XX
XX KW Mycobacterium tuberculosis; Mycobacterium bovis; mycobacterium;
XX detection; BAC vector; bacterial artificial chromosome; tuberculosis;
XX gene; ds.
XX
XX OS Mycobacterium tuberculosis.
XX
XX PN WO9954487-A2.
XX
XX PD 28-OCT-1999.

```


XX PF 16-APR-1999; 99WO-IB00740.
 XX PR 16-APR-1998; 98US-0060756.
 XX PA (INSP) INST PASTEUR.
 XX PI Cole S, Buchrieser-Brosch R, Gordon S, Billault A;
 XX DR WPI; 2000-013262/01.
 XX PT Isolation of polynucleotides from mycobacterial genomes, useful for
 XX detection of Mycobacteria and for combating tuberculosis -
 XX PS Claim 23; Page 62; 161pp; English.
 XX CC The present invention describes a method for isolating a polynucleotide
 CC of interest that is present or is expressed in a genome of a first
 CC mycobacterium strain and that is absent or altered in a genome of a
 CC second mycobacterium strain, which is different from the first strain
 CC using a bacterial artificial chromosome (BAC) vector. Recombinant BAC
 CC vectors, which are preferably immobilised, can be used to detect
 CC mycobacterial nucleic acids (genomic DNA, cDNA or mRNA) in biological
 CC samples. The polynucleotides identified are useful as probes or primers
 CC for detecting a given mycobacterium of interest. By aligning the
 CC polynucleotides contained in the recombinant BAC vectors it is possible
 CC to physically map a polynucleotide of mycobacterial origin in a
 CC biological sample. The methods and vectors from the present invention
 CC are useful in providing information for combating tuberculosis. It is
 CC possible to compare genomes between different strains or species and
 CC their non-pathogenic strains or species counterparts. ABQ62492 to
 CC ABQ63228 and AB881227 to AB881230 represent sequences used in the
 CC exemplification of the present invention.
 XX SQ Sequence 240 BP; 43 A; 70 C; 74 G; 53 T; 0 other;
 Query Match 81.1%; Score 15.4; DB 21; Length 240;
 Best Local Similarity 94.1%; Pred. No. 3.9e+02;
 Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 Qy 3 CGGTGCAGGAGCGTGACA 19
 Db 110 CGGTGCAGGAGCGTGACA 94
 RESULT 12
 AAS66562/c
 ID AAS66562 standard; cDNA; 1202 BP.
 XX AC AAS66562;
 XX DT 13-FEB-2002 (first entry)
 XX DE DNA encoding novel human diagnostic protein #2366.
 XX KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
 XX food supplement; medical imaging; diagnostic; genetic disorder; ss.
 XX OS Homo sapiens.
 XX PN WO200175067-A2.
 XX PD 11-OCT-2001.
 XX PF 30-MAR-2001; 2001WO-US08631.
 XX PR 31-MAR-2000; 2000US-0540217.
 XX PR 23-AUG-2000; 2000US-0649167.
 XX PA (HYSE-) HYSEQ INC.
 XX PI Drmanac RT, Liu C, Tang YT;
 XX DR WPI; 2001-192469/20.
 DR WPI; 2001-639362/73.
 DR P-PSDB; ABG02375.
 XX CC New isolated polynucleotide and encoded polypeptides, useful in
 XX diagnostics, forensics, gene mapping, identification of mutations
 XX responsible for genetic disorders or other traits and to assess
 XX biodiversity -
 XX PS Claim 1; SEQ ID No 2366; 103pp; English.
 XX CC The invention relates to isolated polynucleotide (I) and
 CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
 CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
 CC and gene mapping, and in recombinant production of (II). The
 CC polynucleotides are also used in diagnostics as expressed sequence tags
 CC for identifying expressed genes. (I) is useful in gene therapy techniques
 CC to restore normal activity of (II) or to treat disease states involving
 CC a food supplement. (II) and its binding partners are useful in medical
 CC imaging of sites expressing (II). (I) and (II) are useful for treating
 CC disorders involving aberrant protein expression or biological activity.
 CC The polypeptide and polynucleotide sequences have applications in
 CC diagnostics, forensics, gene mapping, identification of mutations
 CC and to produce other types of data and products dependent on DNA and
 CC amino acid sequences. AAS64197-AAS94564 represent novel human
 CC diagnostic coding sequences of the invention.
 CC Note: The sequence data for this patent did not appear in the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.
 XX SQ Sequence 1202 BP; 290 A; 323 C; 304 G; 285 T; 0 other;
 Query Match 81.1%; Score 15.4; DB 23; Length 1202;
 Best Local Similarity 94.1%; Pred. No. 4.1e+02;
 Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 Qy 2 TCGGTGCAGGAGCGTGAC 18
 Db 75 TGGGTGCAGGAGCGTGAC 59
 RESULT 13
 AAF60970
 ID AAF60970 standard; DNA; 1332 BP.
 XX AC AAF60970;
 XX DT 16-MAY-2001 (first entry)
 XX DE P. putida KT2440-associated DNA ORF00652.
 XX KW Transgenic plant; detection; probe; amplification; vaccine carrier;
 XX microbial production strain; biological remediation; ds.
 XX OS Pseudomonas putida.
 XX PN DE19935088-A1.
 XX PD 01-FEB-2001.
 XX PF 27-JUL-1999; 99DE-1035088.
 XX PR 27-JUL-1999; 99DE-1035088.
 XX PA (TIGR-) TIGR INST GENOMIC RES.
 XX PA (QUIA-) QUIAGEN GMBH.
 XX PA (GBFB) GES BIOTECHNOLOGISCHE FORSCHUNG MBH.
 XX PA (DKFZ-) DKFZ DEUT KREBSFORSCHUNGSZENTRUM.
 XX PA (MEDI-) MEDIZINISCHE HOCHSCHULE HANNOVER.
 XX DR WPI; 2001-192469/20.

XX New DNA sequences specific for *Pseudomonas putida* KT2440, useful as
PT safe genetic engineering host, allow detection in presence of other
PT related bacteria -
XX
PS Claim 1a; Page 17-18; 158pp; German.
XX
XX This invention describes novel DNA sequences (I) for specific detection
CC of *Pseudomonas putida* KT2440. The invention also describes (1)
CC recombinant expression vector containing (1); (2) prokaryotic or
CC eukaryotic cells transformed or transfected with (1) or the vector of
CC (1); (3) production of expression products by culturing cells of (2);
CC (4) expression products, or their fragments, of (1) and synthetic
CC proteins or peptides with the same sequences (A); (5) poly- or
CC mono-clonal antibodies (Ab) that react specifically with (A); (6)
CC hybridoma cells that produce the monoclonal Ab of (5); (7) transgenic
CC plants that contain transformed or transfected cells of (2); (8)
CC detecting KT2440 using a labeled (1) or Ab as probe; and (9) DNA chips
CC carrying one or more (1), (1), and their fragments, are used as probes
CC to detect and isolate full-length cDNAs and/or to amplify such cDNAs by
CC polymerase chain reaction, and for production of transgenic plants. (1),
CC or antibodies that recognize their expression products, are used for
CC detecting the presence of KT2440, particularly in presence of other,
CC even closely related, bacteria. KT2440 is one of the bacteria classified
CC as safe, by the National Institutes of Health, for genetic engineering
CC work, e.g. as microbial production strains, for biological remediation
CC and as vaccine carriers. (1) are exclusive to KT2440 with no significant
CC homology with sequences in other bacteria (specifically the closely
CC related pathogen *P. aeruginosa*). Compared with other 'safe' bacteria, it
CC has greater catabolic activity and better survival in, and adaptation to,
CC the rhizosphere and soil.
XX
XX Sequence 1332 BP; 182 A; 360 C; 486 G; 304 T; 0 other;
SQ

Query Match 81.1%; Score 15.4; DB 22; Length 1332;
Best Local Similarity 94.1%; Pred. No. 4.1e+02;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GTCGGTGCAGGACGTGA 17
|| |||||
DB 1230 GACCGTGCAGGACGTGA 1246

RESULT 14
AAI99683
ID AAI99683 standard; DNA; 4403765 BP.
XX
XX AAI99683;
AC
XX
XX 15-JAN-2002 (first entry)
DT
XX
XX Mycobacterium tuberculosis strain H37Rv genome SEQ ID NO 2.
DE
XX
XX Mycobacterium tuberculosis; strain H37Rv; strain CDC 1551; genome;
KW variation; epidemiology; patient treatment; epidemic monitoring; ds.
KW
XX
XX Mycobacterium tuberculosis.
OS
XX
XX US6294328-B1.
XX
XX 25-SEP-2001.
PD
XX
XX 24-JUN-1998; 98US-0103840.
PF
XX
XX 24-JUN-1998; 98US-0103840.
PR
XX
XX (GENO-) INST GENOMIC RES.
XX
XX Fleischmann RD, White OR, Fraser CM, Venter JC;
XX WPI; 2001-647261/74.
XX
XX Evaluating strain variation of Mycobacterium tuberculosis, comprises

PT determining the nucleotide sequence of the strain at positions in the
PT genome corresponding to positions where *M. tuberculosis* strains CDC
PT 1551 and H37Rv differ -
XX
XX Claim 4; SEQ ID NO 2; 3pp + Sequence Listing; English.
PS
XX
XX The invention relates to evaluating strain variation within and between
CC different populations of the tuberculosis bacterial pathogen,
CC Mycobacterium tuberculosis or related Mycobacterium by determining the
CC nucleotide sequence of the first strain at positions in the complete
CC sequence of the genome that correspond to positions that differ in the
CC nucleotide sequences of *M. tuberculosis* strains CDC 1551 (AAI99683) and
CC H37Rv (AAI99682). The method is useful for evaluating strain variation of
CC *M. tuberculosis* and has valuable application in the fields of
CC tuberculosis genetics, epidemiology, patient treatment and epidemic
CC monitoring
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from USPTO
CC at seqdata.uspto.gov/sequence.html?DocID=6294328B1.
XX
XX Sequence 4403765 BP; 757105 A; 1447799 C; 1441301 G; 757371 T; 189 other;
SQ

Query Match 81.1%; Score 15.4; DB 22; Length 4403765;
Best Local Similarity 94.1%; Pred. No. 3.3e+02;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GTCGGTGCAGGACGTGA 17
|| |||||
DB 1664736 GTTGGTGCAGGACGTGA 1664752

RESULT 15
AAI99682
ID AAI99682 standard; DNA; 4411529 BP.
XX
XX AAI99682;
AC
XX
XX 15-JAN-2002 (first entry)
DT
XX
XX Mycobacterium tuberculosis strain H37Rv genome SEQ ID NO 1.
DE
XX
XX Mycobacterium tuberculosis; strain H37Rv; strain CDC 1551; genome;
KW variation; epidemiology; patient treatment; epidemic monitoring; ds.
KW
XX
XX Mycobacterium tuberculosis.
OS
XX
XX US6294328-B1.
XX
XX 25-SEP-2001.
PD
XX
XX 24-JUN-1998; 98US-0103840.
PF
XX
XX 24-JUN-1998; 98US-0103840.
PR
XX
XX (GENO-) INST GENOMIC RES.
XX
XX Fleischmann RD, White OR, Fraser CM, Venter JC;
XX WPI; 2001-647261/74.
XX
XX Evaluating strain variation of Mycobacterium tuberculosis, comprises
PT determining the nucleotide sequence of the strain at positions in the
PT genome corresponding to positions where *M. tuberculosis* strains CDC
PT 1551 and H37Rv differ -
XX
XX Claim 3; SEQ ID NO 1; 3pp + Sequence Listing; English.
PS
XX
XX The invention relates to evaluating strain variation within and between
CC different populations of the tuberculosis bacterial pathogen,
CC Mycobacterium tuberculosis or related Mycobacterium by determining the
CC nucleotide sequence of the first strain at positions in the complete
CC sequence of the genome that correspond to positions that differ in the
CC nucleotide sequences of *M. tuberculosis* strains CDC 1551 (AAI99683) and
CC H37Rv (AAI99682). The method is useful for evaluating strain variation of
CC *M. tuberculosis* and has valuable application in the fields of
CC tuberculosis genetics, epidemiology, patient treatment and epidemic
CC monitoring

CC H37Rv (AAI99682). The method is useful for evaluating strain variation of
CC M. tuberculosis and has valuable application in the fields of
CC tuberculosis genetics, epidemiology, patient treatment and epidemic
CC monitoring.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from USPTO
CC at seqdata.uspto.gov/sequence.html?DocID=6294328B1.
XX

SQ Sequence 4411529 BP; 758565 A; 1449983 C; 1444602 G; 758379 T; 0 other;

Query Match 81.1%; Score 15.4; DB 22; Length 4411529;
Best Local Similarity 94.1%; Pred. No. 3.3e+02;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 GTCGGTGCAGGACGTGA 17
|| |||||
Db 1664602 GTTGGTGCAGGACGTGA 1664618

Search completed: January 3, 2003, 23:20:41
Job time : 1049.77 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 3, 2003, 22:54:17 ; Search time 252,634 Seconds
(without alignments)
1218.024 Million cell updates/sec

Title: US-09-787-562-2

Perfect score: 19
Sequence: 1 gtcggtgcaggacgtgaca 19

Scoring table: IDENTITY NUC
Gapop 10.0, Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

```

1: em_estba.*
2: em_esthum.*
3: em_estin.*
4: em_estmu.*
5: em_estov.*
6: em_estpl.*
7: em_estro.*
8: em_hc.*
9: gb_estl.*
10: gb_est2.*
11: gb_hc.*
12: gb_est3.*
13: gb_est4.*
14: gb_est5.*
15: em_estfun.*
16: em_estom.*
17: gb_gss.*
18: em_gss_hum.*
19: em_gss_inv.*
20: em_gss_pln.*
21: em_gss_vrt.*
22: em_gss_fun.*
23: em_gss_mam.*
24: em_gss_mus.*
25: em_gss_other.*
26: em_gss_pro.*
27: em_gss_rod.*

```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
c 1	16.4	86.3	567	13	BM140346 WHE0474_9
2	16.4	86.3	872	17	CNS03302
c 3	16.4	86.3	1021	17	AL225515 Tetraodon
c 4	16.4	86.3	1101	17	AL318185 Tetraodon
5	16	84.2	295	9	AL319040 Tetraodon
6	16	84.2	734	13	AI903141 QV-BT022- BI838592 603086272

7	16	84.2	786	13	BI819311
8	16	84.2	792	13	BI821807
9	16	84.2	799	13	BI916645
10	16	84.2	810	13	BI751993
11	16	84.2	819	13	BI829000
12	16	84.2	856	12	BG745622
c 13	16	84.2	884	9	AL537598
14	16	84.2	914	9	AL537599
15	16	84.2	1065	13	BM547868
16	16	84.2	1084	14	BM919043
17	16	84.2	1096	14	BM805234
18	15.8	83.2	392	10	AW436851
19	15.8	83.2	408	10	AW416022
20	15.8	83.2	442	17	AZ247056
c 21	15.8	83.2	448	10	BB839526
c 22	15.8	83.2	449	10	BB839535
23	15.8	83.2	475	12	BG333459
24	15.8	83.2	479	9	AA547770
25	15.8	83.2	531	17	AZ156676
26	15.8	83.2	556	10	AV616722
27	15.8	83.2	568	9	AJ273287
28	15.8	83.2	610	12	BG220636
29	15.8	83.2	784	17	BH400842
30	15.8	83.2	827	13	BI533157
31	15.8	83.2	907	17	CNS01KDC
c 32	15.4	81.1	250	10	BE011583
33	15.4	81.1	280	9	AJ493693
34	15.4	81.1	341	10	AW425923
c 35	15.4	81.1	405	9	AA933175
36	15.4	81.1	425	10	AV613228
37	15.4	81.1	442	10	BE481127
c 38	15.4	81.1	446	10	AW503758
c 39	15.4	81.1	453	10	AW520512
40	15.4	81.1	467	9	AL695680
c 41	15.4	81.1	469	9	AA998465
42	15.4	81.1	499	13	BI468593
43	15.4	81.1	507	13	BM588599
44	15.4	81.1	509	13	BM622212
45	15.4	81.1	519	13	BM610964

ALIGNMENTS

RESULT 1
BM140346/c
LOCUS
DEFINITION
WHE0474_g05_n10zs Wheat Fusarium graminearum infected spike cDNA library Triticum aestivum cDNA clone WHE0474_g05_n10, mRNA sequence.
ACCESSION
BM140346
VERSION
BM140346.1 GI:17151421
KEYWORDS
EST.
SOURCE
bread wheat.
ORGANISM
Triticum aestivum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae; Triticeae; Triticum.
REFERENCE
1 (bases 1 to 567)
AUTHORS
Anderson, O.D., Chao, S., Han, P.S., Heinen, S., Hsia, C.C., Kang, Y., Kruger, W.M., Lazo, G.R., Miller, S., Muehlbauer, G.J., Miller, R., Pritsch, C., Rausch, C.J., Seaton, C.L., Tong, J.C., Vance, C. and Wilson, C.F.
TITLE
The structure and function of the expressed portion of the wheat genomes - Fusarium graminearum infected spike cDNA library
JOURNAL
Unpublished (2001)
COMMENT
Contact: Olin Anderson
US Department of Agriculture, Agriculture Research Service, Pacific West Area, Western Regional Research Center
800 Buchanan Street, Albany, CA 94710, USA
Tel: 5105595773
Fax: 5105595818
Email: oandersn@pw.usda.gov

Mon Jan 6 15:20:23 2003

Sequences have been trimmed to remove vector sequence and low quality sequence with phred score less than 20. No effort was taken to identify ESTs of fungal origin from this library, thus this EST could be of wheat or fungal origin.

Seq primer: Strategene SK primer.

FEATURES

source

1. .567
/organism="Triticum aestivum"
/cultivar="Suma13"
/db_xref="taxon:4565"
/clone="WHE0474_g05.n10"
/clone_lib="wheat Fusarium graminearum infected spike cDNA library"
/tissue_type="Spike"
/dev_stage="Adult plant"
/lab_host="E. coli SOLR"
/note="Vector: Lambda Uni-ZAP XR, excised phagemid pBluescript SK; Site_1: EcoRI; Site_2: XhoI; Plants were grown in the greenhouse. Spikes were sprayed at anthesis with Fusarium graminearum. Total RNA, and poly(A) RNA were prepared and pooled from infected spike at 0, 6, 12, 24, 36 and 48 hours after inoculation, a cDNA library was made, and the cDNA clones were in vivo excised to give pBluescript phagemids in G. Muehlbauer lab at the University of Minnesota (Kruger, W.M., Muehlbauer, G.J., Pritsch, C., Vance, C.). The cDNA library should contain genes of both wheat and fungal pathogen origin. Plasmid DNA preparations and DNA sequencing were performed in the OD Anderson lab (all other authors)."

99 a 173 c 139 g 156 t

Query Match 86.3%; Score 16.4; DB 13; Length 567;

Best Local Similarity 94.4%; Pred. No. 2.9e+03;

Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GTCGGTGCAGGACGTGAC 18

||||| ||||||| |||

Db 523 GTCGGTGCAGGACGGAC 506

RESULT 2

CNS03302

LOCUS

DEFINITION

208b11 of library G from Tetraodon nigroviridis, genomic survey

sequence.

AL225515.1 GI:7884413

GSS: genome survey sequence.

Tetraodon nigroviridis.

Tetraodon nigroviridis

Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;

Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;

Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;

Tetraodontidae; Tetraodon.

1 (bases 1 to 872)

Roest-Crollius, H., Jaillon, O., Dasilva, C., Bouneau, L., Fisher, C.,

Bernot, A., Fizames, C., Wincker, P., Brottier, P., Quetier, F.,

Saurin, W. and Weissenbach, J.

Human gene number estimate provided by genome wide analysis using

Tetraodon nigroviridis DNA sequence

Unpublished

2 (bases 1 to 872)

Roest-Crollius, H., Jaillon, O., Dasilva, C., Fizames, C., Fisher, C.,

Bouneau, L., Billault, A., Quetier, F., Saurin, W., Bernot, A. and

Weissenbach, J.

Characterization and repeat analysis of the compact genome of the

freshwater pufferfish Tetraodon nigroviridis

Unpublished

3 (bases 1 to 872)

Roest-Crollius, H., Jaillon, O., Dasilva, C., Bouneau, L., Fisher, C.,

Bernot, A., Fizames, C., Wincker, P., Brottier, P., Quetier, F.,

Saurin, W. and Weissenbach, J.

Human gene number estimate provided by genome wide analysis using

Tetraodon nigroviridis DNA sequence

Unpublished

2 (bases 1 to 872)

Roest-Crollius, H., Jaillon, O., Dasilva, C., Fizames, C., Fisher, C.,

Bouneau, L., Billault, A., Quetier, F., Saurin, W., Bernot, A. and

Weissenbach, J.

Characterization and repeat analysis of the compact genome of the

freshwater pufferfish Tetraodon nigroviridis

Unpublished

3 (bases 1 to 872)

Roest-Crollius, H., Jaillon, O., Dasilva, C., Bouneau, L., Fisher, C.,

Bernot, A., Fizames, C., Wincker, P., Brottier, P., Quetier, F.,

Saurin, W. and Weissenbach, J.

Human gene number estimate provided by genome wide analysis using

Tetraodon nigroviridis DNA sequence

Unpublished

2 (bases 1 to 872)

Roest-Crollius, H., Jaillon, O., Dasilva, C., Bouneau, L., Fisher, C.,

Bernot, A., Fizames, C., Wincker, P., Brottier, P., Quetier, F.,

Saurin, W. and Weissenbach, J.

Human gene number estimate provided by genome wide analysis using

Tetraodon nigroviridis DNA sequence

JOURNAL

COMMENT

Submitted (12-APR-2000)
This sequence is a single read and was generated as part of a large scale clone-end sequencing project of the Tetraodon nigroviridis genome. For more information, please take a look at <http://www.genoscope.cns.fr/Tetraodon>.

FEATURES

source

1. .872
/organism="Tetraodon nigroviridis"
/db_xref="taxon:99883"
/clone="208E11"
/clone_lib="G"
/note="Genoscope sequence ID : C0AG208AC06SP1-end : PUC-ori"

169 a 269 c 224 g 209 t 1 others

BASE COUNT

ORIGIN

Query Match 86.3%; Score 16.4; DB 17; Length 872;

Best Local Similarity 94.4%; Pred. No. 3.2e+03;

Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GTCGGTGCAGGACGTGAC 18

||||| ||||||| |||

Db 458 GTCGGTGCAGGACGTGAC 475

RESULT 3

CNS05218/c

LOCUS

DEFINITION

031C11 of library A from Tetraodon nigroviridis, genomic survey

sequence.

AL318185

GI:9551069

GSS: genome survey sequence.

Tetraodon nigroviridis.

Tetraodon nigroviridis

Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;

Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;

Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;

Tetraodontidae; Tetraodon.

1 (bases 1 to 1021)

Roest-Crollius, H., Jaillon, O., Dasilva, C., Bouneau, L., Fisher, C.,

Bernot, A., Fizames, C., Wincker, P., Brottier, P., Quetier, F.,

Saurin, W. and Weissenbach, J.

Estimate of human gene number provided by genome-wide analysis

using Tetraodon nigroviridis DNA sequence

Nat. Genet. 25 (2), 235-238 (2000)

20296633

10835645

2 (bases 1 to 1021)

Crollius, H.R., Jaillon, O., Dasilva, C., Ozouf-Costaz, C., Fizames, C.,

Fischer, C., Bouneau, L., Billault, A., Quetier, F., Saurin, W.,

Bernot, A. and Weissenbach, J.

Characterization and repeat analysis of the compact genome of the

freshwater pufferfish Tetraodon nigroviridis

Genome res. 10 (7), 939-949 (2000)

20359837

10899143

3 (bases 1 to 1021)

Genoscope.

Direct Submission

Submitted (12-APR-2000)

This sequence is a single read and was generated as part of a large

scale clone-end sequencing project of the Tetraodon nigroviridis

genome. For more information, please take a look at

<http://www.genoscope.cns.fr/Tetraodon>.

Location/Qualifiers

1. .1021

/organism="Tetraodon nigroviridis"

/db_xref="taxon:99883"

/clone="031C11"

/clone_lib="A"

/note="Genoscope sequence ID : C0AA031AB06A2-end : T3"

FEATURES

source

```

BASE COUNT      309 a      217 c      229 g      255 t      11 others
ORIGIN

Query Match      86.3%; Score 16.4; DB 17; Length 1021;
Best Local Similarity 94.4%; Pred. No. 3.3e+03;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GTCGGTGCAGGACGTGAC 18
    ||| ||||| ||||| |||||
Db 127 GTTGGTGCAGGACGTGAC 110

RESULT 4
CNS0535Z/c
LOCUS
DEFINITION      1101 bp DNA linear GSS 26-JUL-2000
Tetraodon nigroviridis genome survey sequence T3 end of clone
024D18 of library A from Tetraodon nigroviridis, genomic survey
sequence.
ACCESSION      AL319040
VERSION
KEYWORDS
SOURCE
ORGANISM      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
Tetraodontidae; Tetraodon.
1 (bases 1 to 1101)
Roest Crolius,H., Jaillon,O., Dasilva,C., Bouneau,L., Fisher,C.,
Bernot,A., Fizames,C., Wincker,P., Brottier,P., Quetier,F.,
Saurin,W. and Weissenbach,J.
Estimate of human gene number provided by genome-wide analysis
using Tetraodon nigroviridis DNA sequence
Nat. Genet. 25 (2), 235-238 (2000)
20296633
10835645
REFERENCE      2 (bases 1 to 1101)
Crollius,H.R., Jaillon,O., Dasilva,C., Ozouf-Costaz,C., Fizames,C.,
Fischer,C., Bouneau,L., Billault,A., Quetier,P., Saurin,W.,
Bernot,A. and Weissenbach,J.
Characterization and repeat analysis of the compact genome of the
freshwater pufferfish Tetraodon nigroviridis
Genome Res. 10 (7), 939-949 (2000)
20359837
10899143
REFERENCE      3 (bases 1 to 1101)
Genoscope.
Direct Submission
Submitted (12-APR-2000)
This sequence is a single read and was generated as part of a large
scale clone-end sequencing project of the Tetraodon nigroviridis
genome. For more information, please take a look at
http://www.genoscope.cns.fr/Tetraodon.
Location/Qualifiers
1. 1101
/organism="Tetraodon nigroviridis"
/db_xref="taxon:99883"
/clone="024D18"
/clone_lib="A"
/note="Genoscope sequence ID : COAA024DB09A1-end : T3"

BASE COUNT      277 a      268 c      308 g      227 t      21 others
ORIGIN

Query Match      86.3%; Score 16.4; DB 17; Length 1101;
Best Local Similarity 94.4%; Pred. No. 3.4e+03;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GTCGGTGCAGGACGTGAC 18
    ||| ||||| ||||| |||||
Db 94 GTCGCTGCAGGACGTGAC 77

RESULT 5
CNS0535Z/c
LOCUS
DEFINITION      734 bp mRNA linear EST 04-OCT-2001
603086272F1 NIH_MGC_120 Homo sapiens cDNA clone IMAGE:5225344 5',
mRNA sequence.
ACCESSION      BI838592
VERSION
KEYWORDS
SOURCE
ORGANISM      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 734)

```

AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>.
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
 JOURNAL Unpublished (1999)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgapbs-r@mail.nih.gov
 Tissue Procurement: Life Technologies, Inc.
 cDNA Library Preparation: Life Technologies, Inc.
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
 Plate: LLAM1566 row: p column: 17
 High quality sequence start: 5
 High quality sequence stop: 728.

FEATURES
 source

1. .734
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:522344"
 /lab_host="NIH_MGC_120"
 /note="Organ: pooled pancreas and spleen; Vector: pCMV-SPORT6; Site:1: NotI; Site:2: EcoRV (destroyed); RNA source anonymous pool of spleen and pancreas from 28 yo male. Library is oligo-dT primed and directionally cloned (EcoRV site is destroyed upon cloning). Average insert size 1.5 kb, insert size range 1-2.5 kb. Library is normalized and enriched for full-length clones and was constructed by C. Gruber (Invitrogen). Research Genetics tracking code 025. Note: this is a NIH_MGC Library."
 BASE COUNT 144 a 210 c 231 g 149 t
 ORIGIN

Query Match 84.2%; Score 16; DB 13; Length 734;
 Best Local Similarity 100.0%; Pred. No. 4.6e+03;
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GTCGGTCGAGCAGCTG 16
 |||||||||||||||
 Db 412 GTCGGTCGAGCAGCTG 427

RESULT 7
 BI819311 786 bp mRNA linear EST 04-OCT-2001
 LOCUS 603037760F1 NIH_MGC_115 Homo sapiens cDNA clone IMAGE:5178687 5',
 DEFINITION mRNA sequence.
 ACCESSION BI819311
 VERSION BI819311.1 GI:15930861
 KEYWORDS EST.
 SOURCE human.

ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE 1 (bases 1 to 786)
 AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>.
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
 JOURNAL Unpublished (1999)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgapbs-r@mail.nih.gov
 Tissue Procurement: Life Technologies, Inc.
 cDNA Library Preparation: Life Technologies, Inc.
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
 Plate: LLAM1445 row: h column: 16
 High quality sequence stop: 776.
 Location/Qualifiers
 1. 786
 /organism="Homo sapiens"

FEATURES
 source

/db_xref="taxon:9606"
 /clone="IMAGE:5178687"
 /clone_lib="NIH_MGC_115"
 /lab_host="DH10B"
 /note="Organ: pooled brain, lung, testis; Vector: pCMV-SPORT6; Site:1: NotI; Site:2: EcoRV (destroyed); RNA source anonymous pool of 6 male brains, age range 23-27; 1 male lung, age 27; and 1 male testis, age 69. Library is oligo-dT primed and directionally cloned (EcoRV site is destroyed upon cloning). Average insert size 1.8 kb, insert size range 1-3 kb. Library is normalized and enriched for full-length clones and was constructed by C. Gruber (Invitrogen). Research Genetics tracking code 021. Note: this is a NIH_MGC Library."
 BASE COUNT 147 a 249 c 244 g 146 t
 ORIGIN

Query Match 84.2%; Score 16; DB 13; Length 786;
 Best Local Similarity 100.0%; Pred. No. 4.7e+03;
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GTCGGTCGAGCAGCTG 16
 |||||||||||||||
 Db 658 GTCGGTCGAGCAGCTG 673

RESULT 8
 BI821807 792 bp mRNA linear EST 04-OCT-2001
 LOCUS 603035886F1 NIH_MGC_115 Homo sapiens cDNA clone IMAGE:5178686 5',
 DEFINITION mRNA sequence.
 ACCESSION BI821807
 VERSION BI821807.1 GI:15933357
 KEYWORDS EST.
 SOURCE human.

ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE 1 (bases 1 to 792)
 AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>.
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
 JOURNAL Unpublished (1999)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgapbs-r@mail.nih.gov
 Tissue Procurement: Life Technologies, Inc.
 cDNA Library Preparation: Life Technologies, Inc.
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
 Plate: LLAM11440 row: l column: 21
 High quality sequence stop: 789.
 Location/Qualifiers
 1. 792
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:5178686"
 /clone_lib="NIH_MGC_115"
 /lab_host="DH10B"
 /note="Organ: pooled brain, lung, testis; Vector: pCMV-SPORT6; Site:1: NotI; Site:2: EcoRV (destroyed); RNA source anonymous pool of 6 male brains, age range 23-27; 1 male lung, age 27; and 1 male testis, age 69. Library is oligo-dT primed and directionally cloned (EcoRV site is destroyed upon cloning). Average insert size 1.8 kb, insert size range 1-3 kb. Library is normalized and enriched for full-length clones and was constructed by C. Gruber (Invitrogen). Research Genetics tracking code 021. Note: this is a NIH_MGC Library."
 BASE COUNT 146 a 249 c 248 g 149 t
 ORIGIN

FEATURES
 source

1. 792
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:5178686"
 /clone_lib="NIH_MGC_115"
 /lab_host="DH10B"
 /note="Organ: pooled brain, lung, testis; Vector: pCMV-SPORT6; Site:1: NotI; Site:2: EcoRV (destroyed); RNA source anonymous pool of 6 male brains, age range 23-27; 1 male lung, age 27; and 1 male testis, age 69. Library is oligo-dT primed and directionally cloned (EcoRV site is destroyed upon cloning). Average insert size 1.8 kb, insert size range 1-3 kb. Library is normalized and enriched for full-length clones and was constructed by C. Gruber (Invitrogen). Research Genetics tracking code 021. Note: this is a NIH_MGC Library."
 BASE COUNT 146 a 249 c 248 g 149 t
 ORIGIN


```

Query Match      84.2%; Score 16; DB 13; Length 792;
Best Local Similarity 100.0%; Pred. No. 4.7e+03;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GTCGGTGCAGGACGCG 16
    |||||
Db 641 GTCGGTGCAGGACGCG 656

RESULT 9
BI916645
LOCUS
DEFINITION
603179586F1 NIH_MGC_121 Homo sapiens cDNA clone IMAGE:5242893 5',
mRNA sequence.
ACCESSION
BI916645
VERSION
BI916645.1 GI:16180607
KEYWORDS
EST.
SOURCE
human.
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 799)
NIH-MGC http://mgc.nci.nih.gov/.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgabs-r@mail.nih.gov
Tissue Procurement: Life Technologies, Inc.
cDNA Library Preparation: Life Technologies, Inc.
DNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM11482 row: k column: 22
High quality sequence stop: 796.
Location/Qualifiers
1..799
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:5242893"
/clone_lib="NIH_MGC_121"
/lab_host="DH10B"
/note="Organ: brain; Vector: pCMV-SPORT6; Site_1: NotI;
Site_2: EcoRV (destroyed); RNA source anonymous pool of 3
fetal brains, female age 20 weeks, female age 24 weeks,
and male age 26 weeks. Library is oligo-dT primed and
directionally cloned (EcoRV site is destroyed upon
cloning). Average insert size 1.7 kb, insert size range
0.7-3.5 kb. Library is normalized and enriched for
full-length clones and was constructed by C. Gruber
(invitrogen). Research Genetics tracking code 017. Note:
this is a NIH_MGC Library."
BASE COUNT 152 a 254 c 247 g 146 t
ORIGIN

Query Match      84.2%; Score 16; DB 13; Length 799;
Best Local Similarity 100.0%; Pred. No. 4.7e+03;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GTCGGTGCAGGACGCG 16
    |||||
Db 676 GTCGGTGCAGGACGCG 691

RESULT 10
BI751993
LOCUS
DEFINITION
603022103F1 NIH_MGC_114 Homo sapiens cDNA clone IMAGE:5192904 5',
mRNA sequence.
ACCESSION
BI751993
VERSION
BI751993.1 GI:15743571
KEYWORDS
EST.
SOURCE
human.
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 810)
NIH-MGC http://mgc.nci.nih.gov/.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgabs-r@mail.nih.gov
Tissue Procurement: Life Technologies, Inc.
cDNA Library Preparation: Life Technologies, Inc.
DNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM11414 row: k column: 20
High quality sequence stop: 783.
Location/Qualifiers
1..810
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:5192904"
/clone_lib="NIH_MGC_114"
/lab_host="DH10B"
/note="Organ: brain; Vector: pCMV-SPORT6; Site_1: NotI;
Site_2: EcoRV (destroyed); RNA source anonymous pool of 6
male brains, age range 23-27 yo. Library is oligo-dT
primed and directionally cloned (EcoRV site is destroyed
upon cloning). Average insert size 1.5 kb, insert size
range 1-3 kb. Library is normalized and enriched for
full-length clones and was constructed by C. Gruber
(invitrogen). Research Genetics tracking code 019. Note:
this is a NIH_MGC Library."
BASE COUNT 154 a 257 c 251 g 148 t
ORIGIN

Query Match      84.2%; Score 16; DB 13; Length 810;
Best Local Similarity 100.0%; Pred. No. 4.7e+03;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GTCGGTGCAGGACGCG 16
    |||||
Db 692 GTCGGTGCAGGACGCG 707

RESULT 11
BI829000
LOCUS
DEFINITION
603074878F1 NIH_MGC_119 Homo sapiens cDNA clone IMAGE:5166859 5',
mRNA sequence.
ACCESSION
BI829000
VERSION
BI829000.1 GI:15940550
KEYWORDS
EST.
SOURCE
human.
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 819)
NIH-MGC http://mgc.nci.nih.gov/.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgabs-r@mail.nih.gov
Tissue Procurement: Life Technologies, Inc.
cDNA Library Preparation: Life Technologies, Inc.
DNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM11414 row: k column: 20
High quality sequence stop: 783.
Location/Qualifiers
1..819
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:5192904"
/clone_lib="NIH_MGC_114"
/lab_host="DH10B"
/note="Organ: brain; Vector: pCMV-SPORT6; Site_1: NotI;
Site_2: EcoRV (destroyed); RNA source anonymous pool of 6
male brains, age range 23-27 yo. Library is oligo-dT
primed and directionally cloned (EcoRV site is destroyed
upon cloning). Average insert size 1.5 kb, insert size
range 1-3 kb. Library is normalized and enriched for
full-length clones and was constructed by C. Gruber
(invitrogen). Research Genetics tracking code 019. Note:
this is a NIH_MGC Library."
BASE COUNT 154 a 257 c 251 g 148 t
ORIGIN

```

High quality sequence stop: 811.
 Location/Qualifiers
 1. 819
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:5166859"
 /clone_lib="NIH_MGC_119"
 /tissue_type="medulla"
 /lab_host="DH108"
 /note="Organ: Brain; Vector: pCMV-SPORT6; Site:1; NotI; Site:2; EcoRV (destroyed); RNA source normal medulla from anonymous male age 27. Library is oligo-dT primed and directionally cloned (EcoRV site is destroyed upon cloning). Average insert size 1.3 kb, insert size range 0.9-3 kb. Library is normalized and enriched for full-length clones and was constructed by C. Gruber (Invitrogen). Research Genetics tracking code 013. Note: this is a NIH_MGC Library."
 158 a 255 c 251 g 155 t

BASE COUNT 158 a 255 c 251 g 155 t

ORIGIN

Query Match 84.2%; Score 16; DB 13; Length 819;
 Best Local Similarity 100.0%; Pred. No. 4.7e+03;
 Mismatches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GTCGGTGCAGGACGTG 16
 |||
 Db 773 GTCGGTGCAGGACGTG 788
 |||

RESULT 12
 LOCUS BG745622 856 bp mRNA linear EST 15-MAY-2001
 DEFINITION 602723868F1 NIH_MGC_113 Homo sapiens cDNA clone IMAGE:4850344 5', mRNA sequence.
 ACCESSION BG745622
 VERSION BG745622.1 GI:14056275
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 856)
 NIH-MGC http://mgs.nci.nih.gov/
 National Institutes of Health, Mammalian Gene Collection (MGC)
 Unpublished (1999)
 Contact: Robert Strausberg, Ph.D.
 Email: cga@bbs-rmail.nih.gov
 Tissue Procurement: Dr. Mark Watson
 cDNA Library Preparation: Ling Hong/Rubin Laboratory
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
 http://image.llnl.gov
 Plate: LLCM1691 row: g column: 17
 High quality sequence stop: 785.
 Location/Qualifiers
 1. 856
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:4850344"
 /clone_lib="NIH_MGC_113"
 /lab_host="DH108 (phage-resistant)"
 /note="Organ: spleen; Vector: pOTB7; Site:1; XhoI; Site:2; EcoRI; CDNA made by oligo-dT priming, directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGCACGAG(G). Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies). Note: this is a NIH_MGC Library."
 146 a 245 c 299 g 166 t

BASE COUNT 146 a 245 c 299 g 166 t

FEATURES
 source

ORIGIN

Query Match 84.2%; Score 16; DB 12; Length 856;
 Best Local Similarity 100.0%; Pred. No. 4.8e+03;
 Mismatches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GTCGGTGCAGGACGTG 16
 |||
 Db 68 GTCGGTGCAGGACGTG 83
 |||

RESULT 13
 LOCUS AL537598 884 bp mRNA linear EST 13-FEB-2001
 DEFINITION AL537598 LTI_FL013_FBRnl Homo sapiens cDNA clone CS0DF026YK11 3 prime, mRNA sequence.
 ACCESSION AL537598
 VERSION AL537598.1 GI:12801091
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 884)
 Li.W.B., Gruber,C., Jessee,J. and Polayes,D.
 Full-length cDNA libraries and normalization
 Unpublished (2001)
 JOURNAL
 COMMENT Contact: Genoscope
 Genoscope - Centre National de Sequencage
 BP 191 91006 Evry cedex - France
 Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr.
 Location/Qualifiers
 1. 884
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="CS0DF026YK11"
 /clone_lib="LTI_FL013_FBRnl"
 /dev_stage="pooled tissue from post conception fetuses (20 week, 24 week and 26 week)"
 /lab_host="DH108"
 /note="Organ: Fetal brain; Vector: pCMVSPORT 6; 1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-stranded cDNA was digested with Not I and cloned into the Not I and Eco RV sites of the pCMVSPORT 6 vector. Library was constructed by Life Technologies. Contact : Feng Liang Life Technologies, a division of Invitrogen 9800 Medical Center Drive Rockville, Maryland 20850, USA Fax : (1) 301 610 8371 Email : fliang@lifetech.com URL : http://fulllength.invitrogen.com"
 BASE COUNT 185 a 210 c 260 g 211 t 18 others

ORIGIN

Query Match 84.2%; Score 16; DB 9; Length 884;
 Best Local Similarity 100.0%; Pred. No. 4.8e+03;
 Mismatches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GTCGGTGCAGGACGTG 16
 |||
 Db 767 GTCGGTGCAGGACGTG 752
 |||

RESULT 14
 LOCUS AL537599 914 bp mRNA linear EST 13-FEB-2001
 DEFINITION AL537599 LTI_FL013_FBRnl Homo sapiens cDNA clone CS0DF026YK11 5 prime, mRNA sequence.
 ACCESSION AL537599
 VERSION AL537599.1 GI:12801092
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

REFERENCE
 AUTHORS
 TITLE
 JOURNAL
 COMMENT

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 914)
 Li, W.B., Gruber, C., Jessee, J. and Polayres, D.
 Full-length cDNA libraries and normalization
 Unpublished (2001)
 Contact: Genoscope
 Genoscope - Centre National de Sequencage
 BP 191 91006 EVRY cedex - France
 Email: segref@genoscope.cns.fr, Web : www.genoscope.cns.fr.

FEATURES
 source

Location/Qualifiers
 1. .914
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="CS0DF026YK11"
 /dev_stage="LTI_FL013_FBrn1"
 /note="Stage: pooled tissue from post conception fetuses (20 week, 24 week and 26 week)"
 /lab_host="DH108"
 /note="Organ: Fetal brain; Vector: pCMVSPORT 6; 1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-stranded cDNA was digested with Not I and cloned into the Not I and Eco RV sites of the pCMVSPORT 6 vector. Library was constructed by Life Technologies. Contact : Feng Liang Life Technologies, a division of Invitrogen 9800 Medical Center Drive Rockville, Maryland 20850, USA Fax : (1) 301 610 8371 Email : liang@lifetech.com URL : <http://fulllength.invitrogen.com>"

BASE COUNT 170 a 291 c 277 g 176 t
 ORIGIN

Query Match 84.2%; Score 16; DB 9; Length 914;
 Best Local Similarity 100.0%; Pred. No. 4.8e+03;
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GTCGGTGCAGGACGTG 16
 |||||
 Db 659 GTCGGTGCAGGACGTG 674

RESULT 15
 BM547868

LOCUS
 DEFINITION
 ACCESSION
 VERSION
 KEYWORDS
 SOURCE
 ORGANISM

BM547868 1065 bp mRNA linear EST 20-FEB-2002
 AGENCOURT_6507246 NIH_MGC_124 Homo sapiens cDNA clone IMAGE:5727974
 5', mRNA sequence.
 BM547868
 BM547868.1 GI:18781998
 EST.
 human.

REFERENCE
 AUTHORS
 TITLE
 JOURNAL
 COMMENT

Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 1065)
 NIH-MGC <http://mgs.nci.nih.gov/>.
 National Institutes of Health, Mammalian Gene Collection (MGC)
 Unpublished (1999)
 Contact: Robert Strausberg, Ph.D.
 Email: cqapbs@mail.nih.gov
 Tissue Procurement: Invitrogen
 cDNA Library Preparation: Life Technologies, Inc.
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Agencourt Bioscience Corporation
 Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
 Plate: L1AM12722 row: 0 column: 15
 High quality sequence stop: 647.

FEATURES
 source

Location/Qualifiers
 1. .1065
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:5727974"
 /clone_lib="NIH_MGC_124"

/tissue_type="hippocampus"
 /lab_host="DH108"
 /note="Organ: Brain; Vector: pCMV-SPORT6; Site_1: EcorV (destroyed); Site_2: NotI; RNA source male hippocampus, age 27. Library is oligo-dT primed and directionally cloned (EcorV site is destroyed upon cloning). Average insert size 1.4 kb, insert size range 0.9-4 kb. Library is normalized and enriched for full-length clones and was constructed by C. Gruber (Invitrogen). Research Genetics tracking code 012."
 BASE COUNT 197 a 345 c 324 g 199 t
 ORIGIN

Query Match 84.2%; Score 16; DB 13; Length 1065;
 Best Local Similarity 100.0%; Pred. No. 5e+03;
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GTCGGTGCAGGACGTG 16
 |||||
 Db 594 GTCGGTGCAGGACGTG 609

Search completed: January 4, 2003, 01:04:11
 Job time : 256.634 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 3, 2003, 22:53:16 : Search time 6.41325 Seconds
(without alignments)
908.566 Million cell updates/sec

Title: US-09-787-562-2

Perfect score: 19

Sequence: 1 gtcgtgcaggacgtgaca 19

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 441362 seqs, 15338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents_NA.*

1: /cgn2_6/ptodata/2/ina/5A.COMB.seq.*

2: /cgn2_6/ptodata/2/ina/5B.COMB.seq.*

3: /cgn2_6/ptodata/2/ina/6A.COMB.seq.*

4: /cgn2_6/ptodata/2/ina/6B.COMB.seq.*

5: /cgn2_6/ptodata/2/ina/PCTUS.COMB.seq.*

6: /cgn2_6/ptodata/2/ina/backfiles1.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	15.8	83.2	1815	3	US-09-041-545-1
2	15.8	83.2	1815	3	US-09-327-925-1
3	15.4	81.1	240	4	US-09-060-756-197
4	15.4	81.1	4403765	4	US-09-103-840A-2
5	15.4	81.1	4411529	4	US-09-103-840A-1
6	15	78.9	18	2	US-08-693-174-3
7	15	78.9	18	2	US-08-853-236-1
8	15	78.9	18	4	US-09-253-738-3
9	15	78.9	24	2	US-08-693-174-2
10	15	78.9	24	4	US-09-253-738-2
11	15	78.9	41	2	US-08-853-236-2
12	15	78.9	1110	2	US-08-693-174-4
13	15	78.9	1110	4	US-09-253-738-4
14	15	78.9	5382	4	US-09-479-122-21
15	15	78.9	7617	3	US-08-646-538-34
16	15	78.9	7617	4	US-09-503-222-34
17	15	78.9	8387	2	US-08-532-814-1
18	15	78.9	8388	4	US-09-225-509-1
19	15	78.9	9737	4	US-09-479-122-22
20	15	78.9	9737	4	US-09-479-122-23
21	15	78.9	9737	4	US-09-479-122-28
22	15	78.9	9871	4	US-09-479-122-24
23	15	78.9	10060	4	US-09-479-122-25
24	14.8	77.9	1332	4	US-09-134-001C-1374
25	14.8	77.9	6854	4	US-09-194-905-7
26	14.8	77.9	43950	4	US-09-735-934A-3
27	14.8	77.9	152331	3	US-09-128-155-16

28	14.8	77.9	176373	3	US-09-128-155-17
29	14.4	75.8	444	4	US-09-134-001C-343
30	14.4	75.8	733	4	US-09-392-184-15
31	14.4	75.8	826	4	US-08-853-774-2
32	14.4	75.8	826	4	US-08-853-774-3
33	14.4	75.8	826	4	US-08-853-774-4
34	14.4	75.8	1447	4	US-09-484-970B-121
35	14.4	75.8	3691	4	US-09-211-704A-3
36	14.2	74.7	34	1	US-08-357-538-2
37	14.2	74.7	34	1	US-08-357-538-3
38	14.2	74.7	34	1	US-08-476-651-2
39	14.2	74.7	34	1	US-08-476-651-3
40	14.2	74.7	34	5	PCT-US93-10051-2
41	14.2	74.7	34	5	PCT-US93-10051-3
42	14.2	74.7	280	4	US-09-060-756-421
43	14.2	74.7	308	4	US-09-172-108-42
44	14.2	74.7	351	4	US-09-060-756-484
45	14.2	74.7	420	4	US-09-060-756-430

ALIGNMENTS

RESULT 1

US-09-041-545-1
; Sequence 1, Application US/09041545
; Patent No. 6071719
; GENERAL INFORMATION:
; APPLICANT: SATHE, GANESH M.
; APPLICANT: HALSEY, WENDY S.
; APPLICANT: MAO, JOYCE YUE
; TITLE OF INVENTION: ECR 673 : A 7-TRANSMEMBRANE
; TITLE OF INVENTION: G-PROTEIN COUPLED RECEPTOR
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: RATNER & PRESTIA
; STREET: P.O. BOX 980
; CITY: VALLEY FORGE
; STATE: PA
; COUNTRY: USA
; ZIP: 19482
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/041,545
; FILING DATE: 11-MAR-1998
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: PRESTIA, PAUL F
; REGISTRATION NUMBER: 23,031
; REFERENCE/DOCKET NUMBER: GP-70414
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 610-407-0700
; TELEFAX: 610-407-0701
; TELEX: 846169
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1815 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cdna
US-09-041-545-1

Query Match 83.2%; Score 15.8; DB 3; Length 1815;
Best Local Similarity 89.5%; Pred. No. 47;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```

RESULT 4
US-09-103-840A-2
; Sequence 2, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
; APPLICANT: FLEISCHMAN, Robert D.
; APPLICANT: WHITE, Owen R.
; APPLICANT: FRASER, Claire M.
; APPLICANT: VENTER, John C.
; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
; TUBERCULOSIS
; FILE REFERENCE: 24366-20007.00
; CURRENT APPLICATION NUMBER: US/09/103,840A
; CURRENT FILING DATE: 1998-06-24
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 4403765
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis
; FEATURE:
; OTHER INFORMATION: CDC 1551
; OTHER INFORMATION: "n" bases at various positions throughout the sequence
; OTHER INFORMATION: represent a, t, c or g
US-09-103-840A-2

Query Match      81.1%; Score 15.4; DB 4; Length 4403765;
Best Local Similarity 94.1%; Pred. No. 51;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GTCGGTGCAGGACGTGA 17
|| ||||| ||||| |||||
Db 1664736 GTTGGTGCAGGACGTGA 1664752

RESULT 5
US-09-103-840A-1
; Sequence 1, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
; APPLICANT: FLEISCHMAN, Robert D.
; APPLICANT: WHITE, Owen R.
; APPLICANT: FRASER, Claire M.
; APPLICANT: VENTER, John C.
; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
; TUBERCULOSIS
; FILE REFERENCE: 24366-20007.00
; CURRENT APPLICATION NUMBER: US/09/103,840A
; CURRENT FILING DATE: 1998-06-24
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 4411529
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis
; OTHER INFORMATION: H37Rv
US-09-103-840A-1

Query Match      81.1%; Score 15.4; DB 4; Length 4411529;
Best Local Similarity 94.1%; Pred. No. 51;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GTCGGTGCAGGACGTGA 17
|| ||||| ||||| |||||
Db 1664602 GTTGGTGCAGGACGTGA 1664618

RESULT 6
US-08-693-174-3
; Sequence 3, Application US/08693174A
; Patent No. 5942434
; GENERAL INFORMATION:
; APPLICANT: Ratcliffe, Peter John
```

```

QY 1 GTCGGTGCAGGACGTGACA 19
|| ||||| || ||||| || |||||
Db 1169 GTCGGTGCAGGACGTGACA 1187

RESULT 2
US-09-327-925-1
; Sequence 1, Application US/09327925A
; Patent No. 6096868
; GENERAL INFORMATION:
; APPLICANT: SATHE, GANESH M.
; APPLICANT: HALSEY, WENDY S.
; APPLICANT: MAO, JOYCE YUE
; TITLE OF INVENTION: A 7 TRANSMEMBRANE G-PROTEIN
; COUPLED RECEPTOR
; FILE REFERENCE: GP-70414-1
; CURRENT APPLICATION NUMBER: US/09/327,925A
; CURRENT FILING DATE: 1998-06-08
; EARLIER APPLICATION NUMBER: 09/041,545
; EARLIER FILING DATE: 1998-03-11
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 1
; LENGTH: 1815
; TYPE: DNA
; ORGANISM: HOMO SAPIENS
; FEATURE:
; NAME/KEY: UNSURE
; LOCATION: (16)
US-09-327-925-1

Query Match      83.2%; Score 15.8; DB 3; Length 1815;
Best Local Similarity 89.5%; Pred. No. 47;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GTCGGTGCAGGACGTGACA 19
|| ||||| || ||||| || |||||
Db 1169 GTCGGTGCAGGACGTGACA 1187

RESULT 3
US-09-060-756-197/c
; Sequence 197, Application US/09060756
; Patent No. 6183957
; GENERAL INFORMATION:
; APPLICANT: Cole, Stewart
; APPLICANT: Buchrieser-Brosch, Roland
; APPLICANT: Gordon, Stephen
; APPLICANT: Billault, Alain
; TITLE OF INVENTION: METHOD FOR ISOLATING A POLYNUCLEOTIDE OF INTEREST FROM
; THE GENOME OF A MYCOBACTERIUM USING A BAC-BASED DNA
; DETECTION OF MYCOBACTERIA
; FILE REFERENCE: 3495-0169
; CURRENT APPLICATION NUMBER: US/09/060,756
; CURRENT FILING DATE: 1998-04-16
; NUMBER OF SEQ ID NOS: 743
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 197
; LENGTH: 240
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis
US-09-060-756-197

Query Match      81.1%; Score 15.4; DB 4; Length 240;
Best Local Similarity 94.1%; Pred. No. 68;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 3 CGGTGCAGGACGTGACA 19
|| ||||| ||||| |||||
Db 110 CGGTGCAGGACGTGACA 94
```

; APPLICANT: Firth, John David
; APPLICANT: Harris, Adrian Llewellyn
; APPLICANT: Pugh, Christopher William
; APPLICANT: Stratford, Ian James
; TITLE OF INVENTION: Targeting Gene Therapy
; FILE REFERENCE: 08/693174
; CURRENT APPLICATION NUMBER: US/08/693,174A
; CURRENT FILING DATE: 1996-12-12
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 3
; LENGTH: 18
; TYPE: DNA
; ORGANISM: Murinae gen. sp.
US-08-693-174-3

Query Match 78.9%; Score 15; DB 2; Length 18;
Best Local Similarity 100.0%; Pred. No. 98;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 GTGCAGGACGTGACA 19
Db 4 GTGCAGGACGTGACA 18
|||||

RESULT 7
US-08-853-236-1
; Sequence 1, Application US/08853236
; Patent No. 5952226
; GENERAL INFORMATION:
; APPLICANT: Aebischer, Patrick
; APPLICANT: Deglon, Nicole
; APPLICANT: Regulier, Etienne
; APPLICANT: Rinsch, Christopher
; TITLE OF INVENTION: DEVICE AND METHOD FOR DELIVERY OF ERYTHROPOIETIN
; FILE REFERENCE: Modex 004 divisional PGK-1 HRE
; CURRENT APPLICATION NUMBER: US/08/853,236
; CURRENT FILING DATE: 1997-05-09
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 18
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: PGK-1 hypoxia
US-08-853-236-1

Query Match 78.9%; Score 15; DB 2; Length 18;
Best Local Similarity 100.0%; Pred. No. 98;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 GTGCAGGACGTGACA 19
Db 4 GTGCAGGACGTGACA 18
|||||

RESULT 8
US-09-253-738-3
; Sequence 3, Application US/09253738
; Patent No. 6265390
; GENERAL INFORMATION:
; APPLICANT: Ratcliffe, Peter John
; APPLICANT: Firth, John David
; APPLICANT: Harris, Adrian Llewellyn
; APPLICANT: Pugh, Christopher William
; APPLICANT: Stratford, Ian James
; TITLE OF INVENTION: Targeting Gene Therapy
; FILE REFERENCE: 08/693174
; CURRENT APPLICATION NUMBER: US/09/253,738
; CURRENT FILING DATE: 1999-02-22
; NUMBER OF SEQ ID NOS: 5

; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 3
; LENGTH: 18
; TYPE: DNA
; ORGANISM: Murinae gen. sp.
US-09-253-738-3

Query Match 78.9%; Score 15; DB 4; Length 18;
Best Local Similarity 100.0%; Pred. No. 98;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 GTGCAGGACGTGACA 19
Db 4 GTGCAGGACGTGACA 18
|||||

RESULT 9
US-08-693-174-2
; Sequence 2, Application US/08693174A
; Patent No. 5942434
; GENERAL INFORMATION:
; APPLICANT: Ratcliffe, Peter John
; APPLICANT: Firth, John David
; APPLICANT: Harris, Adrian Llewellyn
; APPLICANT: Pugh, Christopher William
; APPLICANT: Stratford, Ian James
; TITLE OF INVENTION: Targeting Gene Therapy
; FILE REFERENCE: 08/693174
; CURRENT APPLICATION NUMBER: US/08/693,174A
; CURRENT FILING DATE: 1996-12-12
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
; LENGTH: 24
; TYPE: DNA
; ORGANISM: Murinae gen. sp.
US-08-693-174-2

Query Match 78.9%; Score 15; DB 2; Length 24;
Best Local Similarity 100.0%; Pred. No. 99;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 GTGCAGGACGTGACA 19
Db 7 GTGCAGGACGTGACA 21
|||||

RESULT 10
US-09-253-738-2
; Sequence 2, Application US/09253738
; Patent No. 6265390
; GENERAL INFORMATION:
; APPLICANT: Ratcliffe, Peter John
; APPLICANT: Firth, John David
; APPLICANT: Harris, Adrian Llewellyn
; APPLICANT: Pugh, Christopher William
; APPLICANT: Stratford, Ian James
; TITLE OF INVENTION: Targeting Gene Therapy
; FILE REFERENCE: 08/693174
; CURRENT APPLICATION NUMBER: US/09/253,738
; CURRENT FILING DATE: 1999-02-22
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
; LENGTH: 24
; TYPE: DNA
; ORGANISM: Murinae gen. sp.
US-09-253-738-2

Query Match 78.9%; Score 15; DB 4; Length 24;
Best Local Similarity 100.0%; Pred. No. 99;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

us-09-787-562-2.rni

Mon Jan 6 15:20:23 2003

```

QY 5 GTGCAGGACGTGACA 19
  |||
Db 7 GTGCAGGACGTGACA 21

RESULT 11
US-08-853-236-2
; Sequence 2, Application US/08853236
; Patent No. 5952226
; GENERAL INFORMATION:
; APPLICANT: Aebischer, Patrick
; APPLICANT: Deglon, Nicole
; APPLICANT: Regulier, Etienne
; APPLICANT: Rinsch, Christopher
; TITLE OF INVENTION: DEVICE AND METHOD FOR DELIVERY OF ERYTHROPOIETIN
; FILE REFERENCE: Modex 004 divisional PKG-1 HRE
; CURRENT APPLICATION NUMBER: US/08/853,236
; CURRENT FILING DATE: 1997-05-09
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
; LENGTH: 41
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: PKG-1 hypoxia
; OTHER INFORMATION: responsive element tandem repeat
US-08-853-236-2

Query Match 78.9%; Score 15; DB 2; Length 41;
Best Local Similarity 100.0%; Pred. No. 1e+02;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 GTGCAGGACGTGACA 19
  |||
Db 27 GTGCAGGACGTGACA 41

RESULT 12
US-08-693-174-4
; Sequence 4, Application US/08693174A
; Patent No. 5942434
; GENERAL INFORMATION:
; APPLICANT: Ratcliffe, Peter John
; APPLICANT: Firth, John David
; APPLICANT: Harris, Adrian Llewellyn
; APPLICANT: Pugh, Christopher William
; APPLICANT: Stratford, Ian James
; TITLE OF INVENTION: Targeting Gene Therapy
; FILE REFERENCE: 08/693174
; CURRENT APPLICATION NUMBER: US/08/693,174A
; CURRENT FILING DATE: 1996-12-12
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 4
; LENGTH: 1110
; TYPE: DNA
; ORGANISM: Murinae gen. sp.
US-08-693-174-4

Query Match 78.9%; Score 15; DB 2; Length 1110;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 GTGCAGGACGTGACA 19
  |||
Db 637 GTGCAGGACGTGACA 651

RESULT 13
US-09-253-738-4
; Sequence 4, Application US/09253738
; Patent No. 6265390
; GENERAL INFORMATION:
; APPLICANT: Ratcliffe, Peter John
; APPLICANT: Firth, John David
; APPLICANT: Harris, Adrian Llewellyn
; APPLICANT: Pugh, Christopher William
; APPLICANT: Stratford, Ian James
; TITLE OF INVENTION: Targeting Gene Therapy
; FILE REFERENCE: 08/693174
; CURRENT APPLICATION NUMBER: US/09/253,738
; CURRENT FILING DATE: 1999-02-22
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 4
; LENGTH: 1110
; TYPE: DNA
; ORGANISM: Murinae gen. sp.
US-09-253-738-4

Query Match 78.9%; Score 15; DB 4; Length 5382;
Best Local Similarity 100.0%; Pred. No. 1.2e+02;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 GTGCAGGACGTGACA 19
  |||
Db 899 GTGCAGGACGTGACA 913

RESULT 14
US-09-479-122-21
; Sequence 21, Application US/09479122
; Patent No. 6410266
; GENERAL INFORMATION:
; APPLICANT: HARRINGTON, JOHN J.
; APPLICANT: SHEREF, BRUCE
; APPLICANT: RUNDLETT, STEPHEN
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR NON-TARGETED ACTIVATION OF
; TITLE OF INVENTION: ENDOGENOUS GENES
; FILE REFERENCE: 0221-0003C
; CURRENT APPLICATION NUMBER: US/09/479,122
; CURRENT FILING DATE: 2000-01-07
; PRIOR APPLICATION NUMBER: 09/276,820
; PRIOR FILING DATE: 1999-03-26
; PRIOR APPLICATION NUMBER: 09/159,643
; PRIOR FILING DATE: 1998-09-24
; PRIOR APPLICATION NUMBER: 08/941,223
; PRIOR FILING DATE: 1997-09-26
; PRIOR APPLICATION NUMBER: 09/263,814
; PRIOR FILING DATE: 1999-03-08
; PRIOR APPLICATION NUMBER: 09/253,022
; PRIOR FILING DATE: 1999-02-19
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 21
; LENGTH: 5382
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (890)
; OTHER INFORMATION: a, c, t, g, other or unknown
; NAME/KEY: modified_base
; LOCATION: (1042)
; OTHER INFORMATION: a, c, t, g, other or unknown
US-09-479-122-21

Query Match 78.9%; Score 15; DB 4; Length 5382;
Best Local Similarity 100.0%; Pred. No. 1.2e+02;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 GTGCAGGACGTGACA 19
  |||
Db 899 GTGCAGGACGTGACA 913

```



```

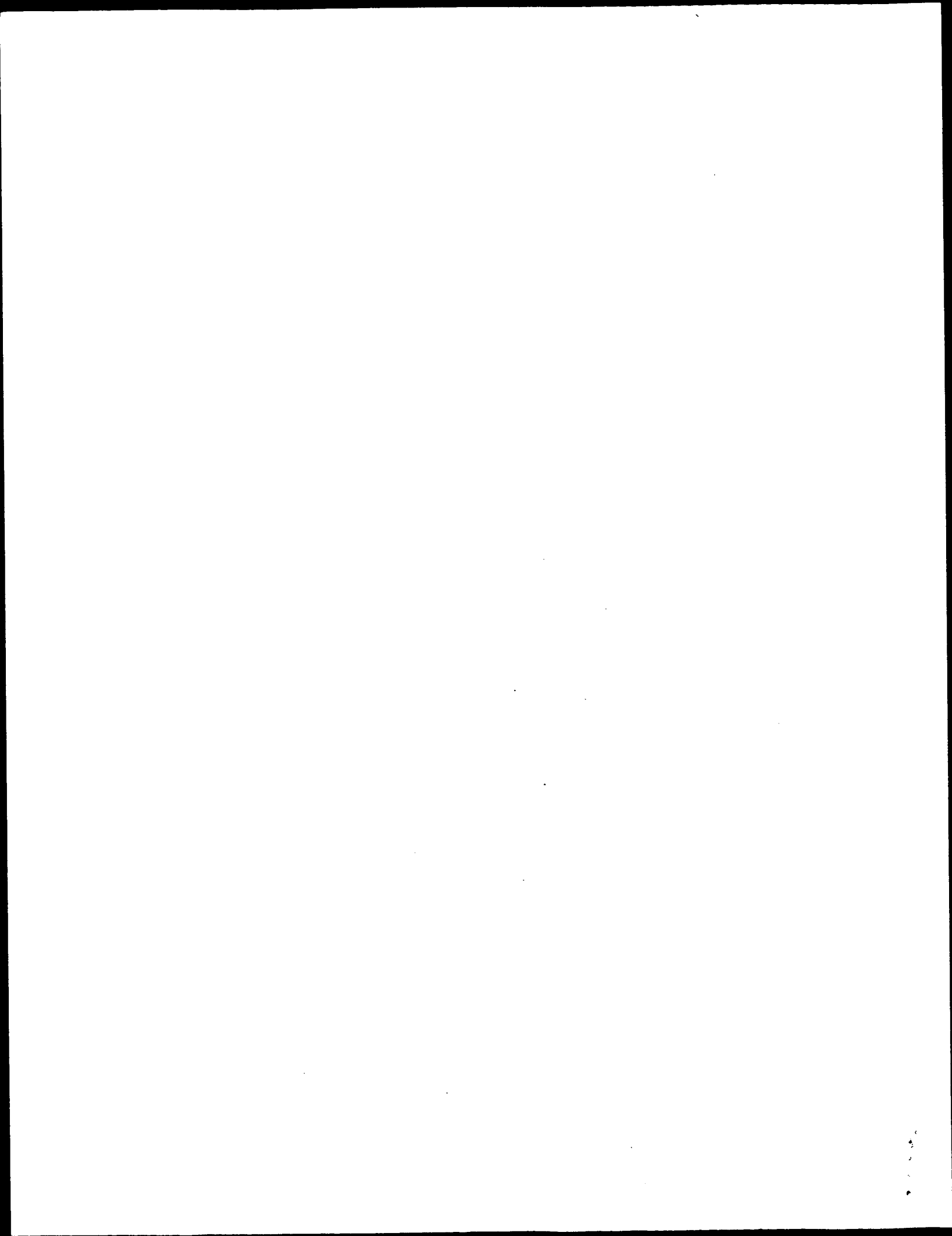
RESULT 15
US-08-646-538-34/c
; Sequence 34, Application US/08646538
; Patent No. 6027881
; GENERAL INFORMATION:
; APPLICANT: Pavlakis, George N.
; APPLICANT: Gaitanaris, George A.
; APPLICANT: Stauber, Roland H.
; APPLICANT: Vournakis, John N.
; TITLE OF INVENTION: Mutant Aequorea victoria Fluorescent
; NUMBER OF SEQUENCES: 37
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Townsend and Townsend and Crew LLP
; STREET: Two Embarcadero Center, 8th Floor
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94111-3834
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC Compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/646,538
; FILING DATE: No. 6027881 yet assigned
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Weber, Kenneth A.
; REGISTRATION NUMBER: 31,677
; REFERENCE/DOCKET NUMBER: 015280-249000
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 576-0200
; TELEFAX: (415) 576-0300
; INFORMATION FOR SEQ ID NO: 34:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 7617 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA
; FEATURE:
; NAME/KEY: -
; LOCATION: 1..7617
; OTHER INFORMATION: /note= "pGen-PKGgfo25RO"
US-08-646-538-34

Query Match 78.98; Score 15; DB 3; Length 7617;
Best Local Similarity 100.0%; Pred. No. 1.2e+02;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 5 GTGCAGGACGTGACA 19
    |||||
Db 4186 GTGCAGGACGTGACA 4172

Search completed: January 4, 2003, 00:09:48
Job time : 976.413 secs

```



us-09-787-562-2.rnpb

Mon Jan 6 15:20:23 2003

```
; CURRENT FILING DATE: 2001-04-11
; PRIOR APPLICATION NUMBER: 09/516,448
; PRIOR FILING DATE: 2000-02-29
; NUMBER OF SEQ ID NOS: 2050
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 1259
; LENGTH: 943
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(943)
; OTHER INFORMATION: n = A,T,C or G
US-09-833-381-1259

Query Match          78.9%; Score 15; DB 10; Length 943;
Best Local Similarity 100.0%; Pred. No. 73;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 GTGCAGGACGTGACA 19
   |||||
Db 153 GTGCAGGACGTGACA 139

RESULT 3
US-10-087-523-1
; Sequence 1, Application US/10087523
; Publication No. US20020197624A1
; GENERAL INFORMATION:
; APPLICANT: Klein, Robert D.
; APPLICANT: Brennan, Thomas J.
; TITLE OF INVENTION: METHODS OF CREATING CONSTRUCTS USEFUL FOR INTRODUCING
; TITLE OF INVENTION: SEQUENCES INTO EMBRYONIC STEM CELLS
; FILE REFERENCE: 376472000200
; CURRENT APPLICATION NUMBER: US/10/087,523
; CURRENT FILING DATE: 2002-02-28
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 09/193,834
; PRIOR FILING DATE: EARLIER FILING DATE: 1998-11-17
; NUMBER OF SEQ ID NOS: 44
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 1
; LENGTH: 4768
; TYPE: DNA
; ORGANISM: plasmid vector
; OTHER INFORMATION:
US-10-087-523-1

Query Match          78.9%; Score 15; DB 9; Length 4768;
Best Local Similarity 100.0%; Pred. No. 83;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 GTGCAGGACGTGACA 19
   |||||
Db 2824 GTGCAGGACGTGACA 2838

RESULT 4
US-09-816-790-1
; Sequence 1, Application US/09816790
; Patent No. US20020022255A1
; GENERAL INFORMATION:
; APPLICANT: Allen, Keith D.
; APPLICANT: Phillips, Russell
; TITLE OF INVENTION: TRANSGENIC MICE CONTAINING
; TITLE OF INVENTION: SULFOTRANSFERASE GENE DISRUPTIONS
; FILE REFERENCE: R-855
; CURRENT APPLICATION NUMBER: US/09/816,790
; CURRENT FILING DATE: 2001-03-22
; PRIOR APPLICATION NUMBER: US 60/191,240
; PRIOR FILING DATE: 2000-03-22
; PRIOR APPLICATION NUMBER: US 60/204,230
; PRIOR FILING DATE: 2000-05-15
; PRIOR APPLICATION NUMBER: US 60/223,173
; PRIOR FILING DATE: 2000-08-07
US-09-816-790-1

Query Match          78.9%; Score 15; DB 10; Length 4768;
Best Local Similarity 100.0%; Pred. No. 83;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 GTGCAGGACGTGACA 19
   |||||
Db 2824 GTGCAGGACGTGACA 2838

RESULT 5
US-09-861-077-1
; Sequence 1, Application US/09861077
; Patent No. US20020023275A1
; GENERAL INFORMATION:
; APPLICANT: Leviten, Michael W.
; TITLE OF INVENTION: TRANSGENIC MICE CONTAINING MATRIX
; TITLE OF INVENTION: METALLOPROTEASE GENE DISRUPTIONS
; FILE REFERENCE: R-15
; CURRENT APPLICATION NUMBER: US/09/861,077
; CURRENT FILING DATE: 2000-05-17
; PRIOR APPLICATION NUMBER: US 60/204,972
; PRIOR FILING DATE: 2000-05-17
; PRIOR APPLICATION NUMBER: US 60/215,394
; PRIOR FILING DATE: 2000-06-29
; NUMBER OF SEQ ID NOS: 21
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 4768
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Phage vector
US-09-861-077-1

Query Match          78.9%; Score 15; DB 10; Length 4768;
Best Local Similarity 100.0%; Pred. No. 83;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 GTGCAGGACGTGACA 19
   |||||
Db 2824 GTGCAGGACGTGACA 2838

RESULT 6
US-09-815-825-1
; Sequence 1, Application US/09815825
; Patent No. US20020026652A1
; GENERAL INFORMATION:
; APPLICANT: Allen, Keith D.
; APPLICANT: Phillips, Russell
; TITLE OF INVENTION: TRANSGENIC MICE CONTAINING CGMP
; TITLE OF INVENTION: PHOSPHODIESTERASE GENE DISRUPTIONS
; FILE REFERENCE: R-849
; CURRENT APPLICATION NUMBER: US/09/815,825
; CURRENT FILING DATE: 2001-03-22
; PRIOR APPLICATION NUMBER: US 60/191,142
; PRIOR FILING DATE: 2000-03-22
; PRIOR APPLICATION NUMBER: US 60/204,227
; PRIOR FILING DATE: 2000-05-15
; PRIOR APPLICATION NUMBER: US 60/216,765
; PRIOR FILING DATE: 2000-07-06
; PRIOR APPLICATION NUMBER: US 60/219,182
; PRIOR FILING DATE: 2000-07-19
US-09-815-825-1

Query Match          78.9%; Score 15; DB 10; Length 4768;
Best Local Similarity 100.0%; Pred. No. 83;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 GTGCAGGACGTGACA 19
   |||||
Db 2824 GTGCAGGACGTGACA 2838
```

```
; NUMBER OF SEQ ID NOS: 21
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
;   LENGTH: 4768
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Phage vector
US-09-815-825-1

Query Match      78.9%; Score 15; DB 10; Length 4768;
Best Local Similarity 100.0%; Pred. No. 83;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 GTGCAGGACGTGACA 19
Db 2824 GTGCAGGACGTGACA 2838

RESULT 7
US-09-815-935-1
; Sequence 1, Application US/09815935
; Patent No. US20020038466A1
; GENERAL INFORMATION:
; APPLICANT: Allen, Keith D.
; TITLE OF INVENTION: TRANSGENIC MICE CONTAINING MAGNESIUM
; FILE REFERENCE: R-723
; CURRENT APPLICATION NUMBER: US/09/815,935
; CURRENT FILING DATE: 2001-03-22
; PRIOR APPLICATION NUMBER: US 60/191,235
; PRIOR FILING DATE: 2000-03-22
; PRIOR APPLICATION NUMBER: US 60/216,249
; PRIOR FILING DATE: 2000-07-06
; NUMBER OF SEQ ID NOS: 21
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
;   LENGTH: 4768
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Phage vector
US-09-815-935-1

Query Match      78.9%; Score 15; DB 10; Length 4768;
Best Local Similarity 100.0%; Pred. No. 83;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 GTGCAGGACGTGACA 19
Db 2824 GTGCAGGACGTGACA 2838

RESULT 8
US-09-815-944-1
; Sequence 1, Application US/09815944
; Patent No. US20020038467A1
; GENERAL INFORMATION:
; APPLICANT: Allen, Keith D.
; APPLICANT: Matthews, William
; APPLICANT: Moore, Mark
; APPLICANT: Phillips, Russell
; TITLE OF INVENTION: TRANSGENIC MICE CONTAINING MELANOCYTE
; FILE REFERENCE: R-654
; CURRENT APPLICATION NUMBER: US/09/815,944
; CURRENT FILING DATE: 2001-03-22
; PRIOR APPLICATION NUMBER: US 60/191,236
; PRIOR FILING DATE: 2000-03-22
; PRIOR APPLICATION NUMBER: US 60/215,214
; PRIOR FILING DATE: 2000-06-29
; PRIOR APPLICATION NUMBER: US 60/218,075
; PRIOR FILING DATE: 2000-07-12

Query Match      78.9%; Score 15; DB 10; Length 4768;
Best Local Similarity 100.0%; Pred. No. 83;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 GTGCAGGACGTGACA 19
Db 2824 GTGCAGGACGTGACA 2838

RESULT 9
US-10-087-523-2
; Sequence 2, Application US/10087523
; Publication No. US20020197624A1
; GENERAL INFORMATION:
; APPLICANT: Klein, Robert D.
; TITLE OF INVENTION: METHODS OF CREATING CONSTRUCTS USEFUL FOR INTRODUCING
; FILE REFERENCE: 37647200200
; CURRENT APPLICATION NUMBER: US/10/087,523
; CURRENT FILING DATE: 2002-02-28
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 09/193,834
; PRIOR FILING DATE: EARLIER FILING DATE: 1998-11-17
; NUMBER OF SEQ ID NOS: 44
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
;   LENGTH: 6355
;   TYPE: DNA
;   ORGANISM: Plasmid vector
US-10-087-523-2

Query Match      78.9%; Score 15; DB 9; Length 6355;
Best Local Similarity 100.0%; Pred. No. 85;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 GTGCAGGACGTGACA 19
Db 4411 GTGCAGGACGTGACA 4425

RESULT 10
US-09-816-790-2
; Sequence 2, Application US/09816790
; Patent No. US20020022255A1
; GENERAL INFORMATION:
; APPLICANT: Allen, Keith D.
; APPLICANT: Phillips, Russell
; TITLE OF INVENTION: TRANSGENIC MICE CONTAINING
; FILE REFERENCE: R-855
; CURRENT APPLICATION NUMBER: US/09/816,790
; CURRENT FILING DATE: 2001-03-22
; PRIOR APPLICATION NUMBER: US 60/191,240
; PRIOR FILING DATE: 2000-03-22
; PRIOR APPLICATION NUMBER: US 60/204,230
; PRIOR FILING DATE: 2000-05-15
; PRIOR APPLICATION NUMBER: US 60/223,173
; PRIOR FILING DATE: 2000-08-07
; NUMBER OF SEQ ID NOS: 21
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
```

Mon Jan 6 15:20:23 2003

```

; LENGTH: 6355
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Phage vector
US-09-815-825-2

Query Match          78.9%; Score 15; DB 10; Length 6355;
Best Local Similarity 100.0%; Pred. No. 85;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 GTGCAGGACGTGACA 19
   |||||
Db 4411 GTGCAGGACGTGACA 4425

RESULT 13
US-09-815-935-2
; Sequence 2, Application US/09815935
; Patent No. US20020038466A1
; GENERAL INFORMATION:
; APPLICANT: Allen, Keith D.
; TITLE OF INVENTION: TRANSGENIC MICE CONTAINING MAGNESIUM
; TITLE OF INVENTION: DEPENDENT PROTEIN PHOSPHATASE GENE DISRUPTIONS
; FILE REFERENCE: R-723
; CURRENT APPLICATION NUMBER: US/09/815,935
; CURRENT FILING DATE: 2001-03-22
; PRIOR APPLICATION NUMBER: US 60/191,235
; PRIOR FILING DATE: 2000-03-22
; PRIOR APPLICATION NUMBER: US 60/216,249
; PRIOR FILING DATE: 2000-07-06
; NUMBER OF SEQ ID NOS: 21
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 6355
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Phage vector
US-09-815-935-2

Query Match          78.9%; Score 15; DB 10; Length 6355;
Best Local Similarity 100.0%; Pred. No. 85;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 GTGCAGGACGTGACA 19
   |||||
Db 4411 GTGCAGGACGTGACA 4425

RESULT 14
US-09-815-944-2
; Sequence 2, Application US/09815944
; Patent No. US20020038467A1
; GENERAL INFORMATION:
; APPLICANT: Allen, Keith D.
; APPLICANT: Matthews, William
; APPLICANT: Moore, Mark
; APPLICANT: Phillips, Russell
; TITLE OF INVENTION: TRANSGENIC MICE CONTAINING MELANOCYTE
; TITLE OF INVENTION: STIMULATING HORMONE RECEPTOR GENE DISRUPTIONS
; FILE REFERENCE: R-654
; CURRENT APPLICATION NUMBER: US/09/815,944
; CURRENT FILING DATE: 2001-03-22
; PRIOR APPLICATION NUMBER: US 60/191,236
; PRIOR FILING DATE: 2000-03-22
; PRIOR APPLICATION NUMBER: US 60/215,214
; PRIOR FILING DATE: 2000-06-29
; PRIOR APPLICATION NUMBER: US 60/218,075
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/219,167
; PRIOR FILING DATE: 2000-07-19
; NUMBER OF SEQ ID NOS: 21
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2

```

; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 6355
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Phage vector
US-09-815-944-2

Query Match 78.9%; Score 15; DB 10; Length 6355;
Best Local Similarity 100.0%; Pred. No. 85;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 GTCCAGGACGTGACA 19
|||||
Db 4411 GTCCAGGACGTGACA 4425

RESULT 15

US-09-987-601-1
; Sequence 1, Application US/09987601
; Patent No. US20020098223A1
; GENERAL INFORMATION:
; APPLICANT: MOULLIER, Philippe
; APPLICANT: DANOS, Olivier
; APPLICANT: HEARD, Jean-Michel
; APPLICANT: FERRY, Nicholas
; TITLE OF INVENTION: BIOCOMPATIBLE IMPLANT FOR THE EXPRESSION AND IN VIVO
; FILE REFERENCE: 0660-0145-0DIV
; CURRENT APPLICATION NUMBER: US/09/987,601
; CURRENT FILING DATE: 2001-11-15
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 09/225,509
; PRIOR FILING DATE: EARLIER FILING DATE: 1999-01-06
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 09/523,814
; PRIOR FILING DATE: EARLIER FILING DATE: 1996-01-19
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: FR 93/04700
; PRIOR FILING DATE: EARLIER FILING DATE: 1993-04-21
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: FR 93/09185
; NUMBER OF SEQ ID NOS: 1
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 8388
; TYPE: DNA
; ORGANISM: mus musculus, Mo-MuLV, and other
US-09-987-601-1

Query Match 78.9%; Score 15; DB 10; Length 8388;
Best Local Similarity 100.0%; Pred. No. 86;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 GTCCAGGACGTGACA 19
|||||
Db 2238 GTCCAGGACGTGACA 2252

Search completed: January 4, 2003, 01:06:07
Job time : 8.41325 secs

GenCore version 5.1.3
 Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 3, 2003, 22:54:41 ; Search time 79.9968 seconds
 (without alignments)
 1281.345 Million cell updates/sec

Title: US-09-787-562-9
 Perfect score: 237
 Sequence: 1 gctagagtcgtcaggacgt.....cgaggcgcctcgccctctg 237

Scoring table: IDENTITY_NUC
 Gapop 10.0 , Gapext 1.0

Searched: 381593 seqs, 216252194 residues
 Total number of hits satisfying chosen parameters: 763186

Minimum DB seq length: 0
 Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
 Maximum Match 100%
 Listing first 45 summaries

Database :

Published Applications NA:*

- 1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:*
- 2: /cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq:*
- 3: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq:*
- 4: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq:*
- 5: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq:*
- 6: /cgn2_6/ptodata/1/pubpna/PCTUS_PUBCOMB.seq:*
- 7: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq:*
- 8: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:*
- 9: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq:*
- 10: /cgn2_6/ptodata/1/pubpna/US09_PUBCOMB.seq:*
- 11: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq:*
- 12: /cgn2_6/ptodata/1/pubpna/US10_PUBCOMB.seq:*
- 13: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq:*
- 14: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Match	Query Length	ID	Description
1	142.4	60.1	309	10	US-09-965-703-64
2	142.4	60.1	633	9	US-09-925-664-6
3	142.4	60.1	847	10	US-09-950-374-2
4	142.4	60.1	850	10	US-09-950-374-1
5	142.4	60.1	4279	10	US-09-956-988A-1
6	142.4	60.1	4565	10	US-09-759-960-7
7	142.4	60.1	5141	10	US-09-924-859A-9
8	142.4	60.1	5332	8	US-08-961-888-40
9	142.4	60.1	5865	12	US-10-098-035-3
10	142.4	60.1	6620	8	US-08-786-531B-3
11	142.4	60.1	6827	10	US-09-982-610-17
12	142.4	60.1	6984	9	US-10-001-189-45
13	142.4	60.1	7127	7	US-08-778-457-46
14	142.4	60.1	7607	10	US-09-982-610-19
15	142.4	60.1	8578	10	US-09-828-825-1
16	142.4	60.1	8578	10	US-09-828-825-3
17	142.4	60.1	8608	10	US-09-828-825-7
18	142.4	60.1	8623	10	US-09-828-825-5
19	142.4	60.1	8629	10	US-09-828-825-15

20 142.4 60.1 8638 10 US-09-828-825-9 Sequence 9, Appli

21 142.4 60.1 8644 10 US-09-828-825-13 Sequence 13, Appl

22 142.4 60.1 8659 10 US-09-828-825-11 Sequence 11, Appl

23 142.4 60.1 8902 10 US-09-729-418A-1 Sequence 1, Appli

24 142.4 60.1 9108 10 US-09-982-610-45 Sequence 45, Appl

25 142.4 60.1 11152 9 US-09-847-101B-12 Sequence 12, Appl

26 142.4 60.1 11600 9 US-09-847-101B-35 Sequence 35, Appl

27 142.4 60.1 14455 9 US-09-847-101B-15 Sequence 15, Appl

28 142.4 60.1 32480 9 US-09-847-101B-23 Sequence 23, Appl

c 29 141.8 59.8 9077 10 US-09-734-300-1 Sequence 1, Appli

c 30 141.8 59.8 9077 10 US-09-734-300-3 Sequence 3, Appli

31 140.8 59.4 259 10 US-09-919-580-315 Sequence 315, App

32 140.8 59.4 259 10 US-09-919-580-621 Sequence 621, App

33 140.8 59.4 4639 10 US-09-804-481-1 Sequence 1, Appli

34 140.8 59.4 5070 10 US-09-795-006A-41 Sequence 41, Appl

35 140.8 59.4 5432 10 US-09-794-975-9 Sequence 9, Appli

36 140.8 59.4 5446 9 US-09-559-874-5 Sequence 5, Appli

37 140.8 59.4 5446 10 US-09-844-645-3 Sequence 3, Appli

38 140.8 59.4 5458 10 US-09-912-436-11 Sequence 11, Appl

39 140.8 59.4 5458 10 US-09-912-436-12 Sequence 12, Appl

40 140.8 59.4 5614 10 US-09-912-436-7 Sequence 7, Appli

41 140.8 59.4 5614 10 US-09-912-436-8 Sequence 8, Appli

42 140.8 59.4 5651 10 US-09-780-933-6 Sequence 6, Appli

43 140.8 59.4 5695 10 US-09-912-436-9 Sequence 9, Appli

44 140.8 59.4 5695 10 US-09-912-436-10 Sequence 10, Appli

45 140.8 59.4 5864 9 US-09-971-980-1 Sequence 1, Appli

ALIGNMENTS

RESULT 1

US-09-965-703-64
 ; Sequence 64, Application US/09965703
 ; Patent No. US20020119521A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Rohm and Haas Company
 ; APPLICANT: Palli, Subba Reddy
 ; APPLICANT: Kapitskaya, Marianna Zinovjevna
 ; APPLICANT: Cress, Dean Ervin
 ; TITLE OF INVENTION: No. US20020119521A1el Ecdysone Receptor-Based Inducible Gene E
 ; FILE REFERENCE: A01020B
 ; CURRENT APPLICATION NUMBER: US/09/965,703
 ; CURRENT FILING DATE: 2001-09-26
 ; PRIOR APPLICATION NUMBER: 60/191,355
 ; PRIOR FILING DATE: 2000-03-22
 ; PRIOR APPLICATION NUMBER: 60/269,799
 ; PRIOR FILING DATE: 2001-02-20
 ; PRIOR APPLICATION NUMBER: PCT/US01/09050
 ; PRIOR FILING DATE: 2001-03-21
 ; NUMBER OF SEQ ID NOS: 75
 ; SOFTWARE: PatentIn version 3.1
 ; SEQ ID NO 64
 ; LENGTH: 309
 ; TYPE: DNA
 ; ORGANISM: Simian virus 40
 ; FEATURE:
 ; NAME/KEY: misc.feature
 ; OTHER INFORMATION: No. US20020119521A1el Sequence
 US-09-965-703-64

Query Match 60.1%; Score 142.4; DB 10; Length 309;
 Best Local Similarity 99.3%; Pred. No. 6.2e-35;
 Matches 143; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Oy 94 TGCATCTCAATATGTCAGCACCATAGTCGCGCCCTTAACTCGCGCCATCCGCGCCCTAA 153
 |||||
 Db 121 TGCATCTCAATATGTCAGCACCATAGTCGCGCCCTTAACTCGCGCCATCCGCGCCCTAA 180
 |||||

Oy 154 CTCGCGCCAGTTCGCGCCATCTCCGCGCCCATCGCTCACTAAATTTTATTTATTTATTCGAC 213
 |||||
 Db 181 CTCGCGCCAGTTCGCGCCATCTCCGCGCCCATGGCTGACTAAATTTTATTTATTTATTCGAC 240
 |||||

QY 214 AGGCGGAGGCGGCTCGGCTCTG 237
 Db 241 AGGCGGAGGCGGCTCGGCTCTG 264

RESULT 2
 US-09-925-664-6
 ; Sequence 6, Application US/099525664
 ; Patent No. US2002016006A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Denney, Jr., Dan W.
 ; TITLE OF INVENTION: Methods of Treating Lymphoma and Leukemia
 ; FILE REFERENCE: GENOTOPE-06499
 ; CURRENT APPLICATION NUMBER: US/09/925,664
 ; CURRENT FILING DATE: 2001-08-09
 ; PRIOR APPLICATION NUMBER: 09/370,453
 ; PRIOR FILING DATE: 1999-08-09
 ; PRIOR APPLICATION NUMBER: 08/644,664
 ; PRIOR FILING DATE: 1996-05-01
 ; PRIOR APPLICATION NUMBER: 08/761,277
 ; PRIOR FILING DATE: 1996-12-06
 ; NUMBER OF SEQ ID NOS: 80
 ; SOFTWARE: PatentIn version 3.1
 ; SEQ ID NO 6
 ; LENGTH: 633
 ; TYPE: DNA
 ; ORGANISM: SR alpha promoter
 US-09-925-664-6

Query Match 60.1%; Score 142.4; DB 9; Length 633;
 Best Local Similarity 99.3%; Pred. No. 7.7e-35;
 Matches 143; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 94 TGCATCTCAATTAGTCAGCAACCATAGTCCGCGCCCTAACTCCGCCATCCGCCCTAA 153
 Db 151 TGCATCTCAATTAGTCAGCAACCATAGTCCGCGCCCTAACTCCGCCATCCGCCCTAA 210
 QY 154 CTCGCCCATGTCGCCCATCTCCGCCCATCGCTGACTAATTTTTTTTATGTCAG 213
 Db 211 CTCGCCCATGTCGCCCATCTCCGCCCATCGCTGACTAATTTTTTTTATGTCAG 270

QY 214 AGGCGGAGGCGGCTCGGCTCTG 237
 Db 271 AGGCGGAGGCGGCTCGGCTCTG 294

RESULT 3
 US-09-950-374-2
 ; Sequence 2, Application US/09950374
 ; Patent No. US20020141981A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Lawman, Patricia
 ; TITLE OF INVENTION: Materials and Methods for Treating Oncological Disease
 ; FILE REFERENCE: MOR-200XCD1
 ; CURRENT APPLICATION NUMBER: US/09/950,374
 ; CURRENT FILING DATE: 2001-09-10
 ; PRIOR APPLICATION NUMBER: US 09/394,226
 ; PRIOR FILING DATE: 1999-09-13
 ; PRIOR APPLICATION NUMBER: PCT/US99/00787
 ; PRIOR FILING DATE: 1999-01-14
 ; PRIOR APPLICATION NUMBER: US 60/071,497
 ; PRIOR FILING DATE: 1998-01-14
 ; NUMBER OF SEQ ID NOS: 6
 ; SEQ ID NO 2
 ; LENGTH: 847
 ; TYPE: DNA
 ; ORGANISM: Streptococcus
 US-09-950-374-2

Query Match 60.1%; Score 142.4; DB 10; Length 847;
 Best Local Similarity 99.3%; Pred. No. 8.5e-35;
 Matches 143; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 94 TGCATCTCAATTAGTCAGCAACCATAGTCCGCGCCCTAACTCCGCCATCCGCCCTAA 153
 Db 143 TGCATCTCAATTAGTCAGCAACCATAGTCCGCGCCCTAACTCCGCCATCCGCCCTAA 202
 QY 154 CTCGCCCATGTCGCCCATCTCCGCCCATCGCTGACTAATTTTTTTTATGTCAG 213
 Db 203 CTCGCCCATGTCGCCCATCTCCGCCCATCGCTGACTAATTTTTTTTATGTCAG 262

QY 214 AGGCGGAGGCGGCTCGGCTCTG 237
 Db 263 AGGCGGAGGCGGCTCGGCTCTG 286

RESULT 4
 US-09-950-374-1
 ; Sequence 1, Application US/09950374
 ; Patent No. US20020141981A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Lawman, Patricia
 ; TITLE OF INVENTION: Materials and Methods for Treating Oncological Disease
 ; FILE REFERENCE: MOR-200XCD1
 ; CURRENT APPLICATION NUMBER: US/09/950,374
 ; CURRENT FILING DATE: 2001-09-10
 ; PRIOR APPLICATION NUMBER: US 09/394,226
 ; PRIOR FILING DATE: 1999-09-13
 ; PRIOR APPLICATION NUMBER: PCT/US99/00787
 ; PRIOR FILING DATE: 1999-01-14
 ; PRIOR APPLICATION NUMBER: US 60/071,497
 ; PRIOR FILING DATE: 1998-01-14
 ; NUMBER OF SEQ ID NOS: 6
 ; SEQ ID NO 1
 ; LENGTH: 850
 ; TYPE: DNA
 ; ORGANISM: Streptococcus
 US-09-950-374-1

Query Match 60.1%; Score 142.4; DB 10; Length 850;
 Best Local Similarity 99.3%; Pred. No. 8.5e-35;
 Matches 143; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 94 TGCATCTCAATTAGTCAGCAACCATAGTCCGCGCCCTAACTCCGCCATCCGCCCTAA 153
 Db 143 TGCATCTCAATTAGTCAGCAACCATAGTCCGCGCCCTAACTCCGCCATCCGCCCTAA 202
 QY 154 CTCGCCCATGTCGCCCATCTCCGCCCATCGCTGACTAATTTTTTTTATGTCAG 213
 Db 203 CTCGCCCATGTCGCCCATCTCCGCCCATCGCTGACTAATTTTTTTTATGTCAG 262

QY 214 AGGCGGAGGCGGCTCGGCTCTG 237
 Db 263 AGGCGGAGGCGGCTCGGCTCTG 286

RESULT 5
 US-09-956-998A-1
 ; Sequence 1, Application US/09956998A
 ; Patent No. US20020082236A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Black Jr., Charles A.
 ; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR ACTIVATING
 ; TITLE OF INVENTION: GENES OF INTEREST
 ; FILE REFERENCE: 5722-2(35722/191928)
 ; CURRENT APPLICATION NUMBER: US/09/956,998A
 ; CURRENT FILING DATE: 2001-09-20
 ; PRIOR APPLICATION NUMBER: 09/446,402
 ; PRIOR FILING DATE: 1999-12-20
 ; PRIOR APPLICATION NUMBER: 60/050,772
 ; PRIOR FILING DATE: 1997-06-25
 ; NUMBER OF SEQ ID NOS: 19
 ; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ ID NO 1

```
; LENGTH: 4279
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Recombinant Molecule containing multiple cloning
; OTHER INFORMATION: site, kozak sequence, LacZ gene.
; NAME/KEY: misc.feature
; LOCATION: (1)...(64)
; OTHER INFORMATION: Multiple cloning site
; NAME/KEY: misc.feature
; LOCATION: (65)...(79)
; OTHER INFORMATION: Consensus sequence for the "Kozak sequence"
; OTHER INFORMATION: (translation initiation)
; NAME/KEY: prim.transcript
; LOCATION: (80)...(4279)
; OTHER INFORMATION: Beta galactosidase
US-09-956-998A-1

Query Match
Best Local Similarity 60.1%; Score 142.4; DB 10; Length 4279;
Matches 143; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 94 TGCATCTCAATTAGTCAGCAACCATAGTCGCGCCCTAACTCCGCCCATCCCGCCCTAA 153
Db 296 TGCATCTCAATTAGTCAGCAACCATAGTCGCGCCCTAACTCCGCCCATCCCGCCCTAA 355

QY 154 CTCGCCCCAGTTCGCCCATCTCTCCGCCCATCGCTGACTAAATTTTTTATTTATGCAG 213
Db 356 CTCGCCCCAGTTCGCCCATCTCTCCGCCCATCGCTGACTAAATTTTTTATTTATGCAG 415

QY 214 AGCGCGAGCGCGCTCGGCCTCTG 237
Db 416 AGCGCGAGCGCGCTCGGCCTCTG 439

RESULT 6
US-09-759-960-7
; Sequence 7, Application us/09759960
; Patent No. US2001000639A1
; GENERAL INFORMATION:
; APPLICANT: Urban, Robert G.
; APPLICANT: Chicx, Roman M.
; APPLICANT: Collins, Edward J.
; APPLICANT: Hedley, Mary Lynn
; TITLE OF INVENTION: IMMUNOGENIC PEPTIDES FROM THE HPV E7
; TITLE OF INVENTION: PROTEIN
; NUMBER OF SEQUENCES: 33
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson, P.C.
; STREET: 225 Franklin Street
; CITY: Boston
; STATE: MA
; COUNTRY: US
; ZIP: 02110-2804
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: Windows95
; SOFTWARE: FASTSEQ for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/759,960
; FILING DATE:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 09/169,425
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Fraser, Janis K.
; REGISTRATION NUMBER: 34,819
; REFERENCE/DOCKET NUMBER: 08191/004002
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617-542-5070
; TELEFAX: 617-543-8906
; TELEX: 200154
```

```
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4665 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA
US-09-759-960-7

Query Match
Best Local Similarity 60.1%; Score 142.4; DB 10; Length 4665;
Matches 143; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 94 TGCATCTCAATTAGTCAGCAACCATAGTCGCGCCCTAACTCCGCCCATCCCGCCCTAA 153
Db 288 TGCATCTCAATTAGTCAGCAACCATAGTCGCGCCCTAACTCCGCCCATCCCGCCCTAA 347

QY 154 CTCGCCCCAGTTCGCCCATCTCTCCGCCCATCGCTGACTAAATTTTTTATTTATGCAG 213
Db 348 CTCGCCCCAGTTCGCCCATCTCTCCGCCCATCGCTGACTAAATTTTTTATTTATGCAG 407

QY 214 AGCGCGAGCGCGCTCGGCCTCTG 237
Db 408 AGCGCGAGCGCGCTCGGCCTCTG 431

RESULT 7
US-09-924-859A-9
; Sequence 9, Application us/09924859A
; Patent No. US20020137113A1
; GENERAL INFORMATION:
; APPLICANT: Godowski, Paul J.
; APPLICANT: Mark, Melanie R.
; APPLICANT: Sadick, Michael D.
; APPLICANT: Shelton, David L.
; APPLICANT: Wong, Wai Lee Tan
; TITLE OF INVENTION: KINASE RECEPTOR ACTIVATION ASSAY
; FILE REFERENCE: P0854C1P2C1
; CURRENT APPLICATION NUMBER: US/09/924,859A
; CURRENT FILING DATE: 2001-08-08
; PRIOR APPLICATION NUMBER: US/09/417,381
; PRIOR FILING DATE: 1999-10-13
; NUMBER OF SEQ ID NOS: 11
; SEQ ID NO 9
; LENGTH: 5141
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Plasmid
US-09-924-859A-9
```

```
Query Match
Best Local Similarity 60.1%; Score 142.4; DB 10; Length 5141;
Matches 143; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 94 TGCATCTCAATTAGTCAGCAACCATAGTCGCGCCCTAACTCCGCCCATCCCGCCCTAA 153
Db 188 TGCATCTCAATTAGTCAGCAACCATAGTCGCGCCCTAACTCCGCCCATCCCGCCCTAA 247

QY 154 CTCGCCCCAGTTCGCCCATCTCTCCGCCCATCGCTGACTAAATTTTTTATTTATGCAG 213
Db 248 CTCGCCCCAGTTCGCCCATCTCTCCGCCCATCGCTGACTAAATTTTTTATTTATGCAG 307

QY 214 AGCGCGAGCGCGCTCGGCCTCTG 237
Db 308 AGCGCGAGCGCGCTCGGCCTCTG 331
```

```
RESULT 8
US-08-961-888-40
; Sequence 40, Application US/08961888
; Patent No. US20010016351A1
; GENERAL INFORMATION:
```

```

; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 5865
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Retroviral vector derived from Moloney Murine
; OTHER INFORMATION: Leukemia Virus
US-10-098-035-3

Query Match          60.1%; Score 142.4; DB 12; Length 5865;
Best Local Similarity 99.3%; Pred. No. 1.5e-34;
Matches 143; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 94 TGCATCTCAATTAGTCAGCAACCATAGTCCGCGCCCTAACTCCGCCCATCCGCCCTAA 153
      |||||||
Db 1641 TGCATCTCAATTAGTCAGCAACCATAGTCCGCGCCCTAACTCCGCCCATCCGCCCTAA 1700

QY 154 CTCGCCCAAGTTCGCCCATCTCTCGCCCATCGCTGACTAATTTTATTATTCGAG 213
      |||||||
Db 1701 CTCGCCCAAGTTCGCCCATCTCTCGCCCATCGCTGACTAATTTTATTATTCGAG 1760

QY 214 AGCGCGAGCGCGCTCGGCTCTG 237
      |||||||
Db 1761 AGCGCGAGCGCGCTCGGCTCTG 1784

RESULT 10
US-08-786-531B-3
; Sequence 3, Application US/08786531B
; Patent No. US20020015979A1
; GENERAL INFORMATION:
; APPLICANT: Link, Charles J.
; APPLICANT: Levy, John P.
; APPLICANT: Wang, Suming
; APPLICANT: Seregina, Tatiana
; TITLE OF INVENTION: Vehicles for Stable Transfer of Green
; TITLE OF INVENTION: Fluorescent Protein Gene and Methods of Use for Same
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Zarley, McKee, Thomte, Voorhees & Sease
; STREET: 801 Grand Suite 3200
; CITY: Des Moines
; STATE: Iowa
; COUNTRY: United States
; ZIP: 50309
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/786,531B
; FILING DATE: 21-JAN-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/10371
; FILING DATE: 22-JAN-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Nebel, Heidi S.
; REGISTRATION NUMBER: 37,719
; REFERENCE/DOCKET NUMBER: Hgt1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 515-288-3667
; TELEFAX: 515-288-1338
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 6620 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: other nucleic acid
; HYPOTHETICAL: NO

```

```

; APPLICANT: Padgett, Kerstien
; APPLICANT: Sorge, Joseph
; TITLE OF INVENTION: NO. US20010016351alel Vector For Gene Expression
; TITLE OF INVENTION: In Prokaryotic And Eukaryotic Systems
; NUMBER OF SEQUENCES: 40
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Howrey & Simon
; STREET: 1299 Pennsylvania Avenue, N.W.
; CITY: Washington
; STATE: DC
; COUNTRY: USA
; ZIP: 20004-2402
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSEQ for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/961,888
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Auerbach, Jeffrey I
; REGISTRATION NUMBER: 32,680
; REFERENCE/DOCKET NUMBER:
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202 383-7451
; TELEFAX: 202 383-6610
; TELEX:
; INFORMATION FOR SEQ ID NO: 40:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 5532 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: circular
US-08-961-888-40

Query Match          60.1%; Score 142.4; DB 8; Length 5532;
Best Local Similarity 99.3%; Pred. No. 1.5e-34;
Matches 143; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 94 TGCATCTCAATTAGTCAGCAACCATAGTCCGCGCCCTAACTCCGCCCATCCGCCCTAA 153
      |||||||
Db 288 TGCATCTCAATTAGTCAGCAACCATAGTCCGCGCCCTAACTCCGCCCATCCGCCCTAA 347

QY 154 CTCGCCCAAGTTCGCCCATCTCTCGCCCATCGCTGACTAATTTTATTATTCGAG 213
      |||||||
Db 348 CTCGCCCAAGTTCGCCCATCTCTCGCCCATCGCTGACTAATTTTATTATTCGAG 407

QY 214 AGCGCGAGCGCGCTCGGCTCTG 237
      |||||||
Db 408 AGCGCGAGCGCGCTCGGCTCTG 431

RESULT 9
US-10-098-035-3
; Sequence 3, Application US/10098035
; Patent No. US20020141983A1
; GENERAL INFORMATION:
; APPLICANT: University of Southern California
; APPLICANT: Weiner, Leslie P.
; APPLICANT: McMillan, Minnie
; TITLE OF INVENTION: CONSTRUCTION AND USE OF GENES ENCODING
; TITLE OF INVENTION: PATHOGENIC EPITOPES FOR TREATMENT OF AUTOIMMUNE DISEASE
; FILE REFERENCE: 13761-703-00 US
; CURRENT APPLICATION NUMBER: US/10/098,035
; CURRENT FILING DATE: 2002-03-14
; PRIOR APPLICATION NUMBER: US/08/654,737
; PRIOR FILING DATE: 1996-05-29
; NUMBER OF SEQ ID NOS: 12

```

ANTI-SENSE: NO
US-08-786-531B-3

Query Match 60.1%; Score 142.4; DB 8; Length 6620;
Best Local Similarity 99.3%; Pred. No. 1.6e-34;
Matches 143; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 94 TGCATCTCAATTAGTCAGCAACCATAGTCCCGCCCTAACTCCGCCCATCCGCCCTAA 153
DB 2556 TGCATCTCAATTAGTCAGCAACCATAGTCCCGCCCTAACTCCGCCCATCCGCCCTAA 2615
QY 154 CTCGCCCGAGTTCGCCCGCATTCCTCGGCCCATCGCTGACTAATTTTTTTTATTCGAG 213
DB 2616 CTCGCCCGAGTTCGCCCGCATTCCTCGGCCCATCGCTGACTAATTTTTTTTATTCGAG 2675
QY 214 AGCGCGAGCGCCCTCGGCCCTCTG 237
DB 2676 AGCGCGAGCGCCCTCGGCCCTCTG 2699

RESULT 11
US-09-982-610-17
; Sequence 17, Application US/09982610
; Patent No. US20020146420A1
; GENERAL INFORMATION:
; APPLICANT: Genentech, Inc.
; BENNETT, BRIAN D.
; GOEDDEL, DAVID
; LEE, JAMES M.
; MATTHEWS, WILLIAM
; TSAI, SIAO PING
; WOOD, WILLIAM I.
TITLE OF INVENTION: PROTEIN TYROSINE KINASE AGONIST ANTIBODIES
NUMBER OF SEQUENCES: 45
CORRESPONDENCE ADDRESS:
ADDRESSEE: Genentech, Inc.
STREET: 460 Point San Bruno Blvd
CITY: South San Francisco
STATE: California
COUNTRY: USA
ZIP: 94080

COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: WinPatIn (Genentech)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/982,610
FILING DATE: 17-Oct-2001
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/446,648
FILING DATE: 1996-MAY-23
APPLICATION NUMBER: 08/222616
FILING DATE: 04-APR-1994
ATTORNEY/AGENT INFORMATION:
NAME: Lee, Wendy M.
REGISTRATION NUMBER: 40,378
REFERENCE/DOCKET NUMBER: P0821P3PCT
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415/225-1994
TELEFAX: 415/952-9881
TELEX: 910/371-7168
INFORMATION FOR SEQ ID NO: 17:
SEQUENCE CHARACTERISTICS:
LENGTH: 6827 base pairs
TYPE: Nucleic Acid
STRANDEDNESS: Single
TOPOLOGY: Linear
SEQUENCE DESCRIPTION: SEQ ID NO: 17:

Query Match 60.1%; Score 142.4; DB 10; Length 6827;

Best Local Similarity 99.3%; Pred. No. 1.6e-34;
Matches 143; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 94 TGCATCTCAATTAGTCAGCAACCATAGTCCCGCCCTAACTCCGCCCATCCGCCCTAA 153
DB 3453 TGCATCTCAATTAGTCAGCAACCATAGTCCCGCCCTAACTCCGCCCATCCGCCCTAA 3512
QY 154 CTCGCCCGAGTTCGCCCGCATTCCTCGGCCCATCGCTGACTAATTTTTTTTATTCGAG 213
DB 3513 CTCGCCCGAGTTCGCCCGCATTCCTCGGCCCATCGCTGACTAATTTTTTTTATTCGAG 3572
QY 214 AGCGCGAGCGCCCTCGGCCCTCTG 237
DB 3573 AGCGCGAGCGCCCTCGGCCCTCTG 3596

RESULT 12
US-10-001-189-45
; Sequence 45, Application US/10001189
; Patent No. US20020173634A1
; GENERAL INFORMATION:
; APPLICANT: FRASER JR., MALCOLM J.
; APPLICANT: LI, XU
; APPLICANT: BEAM, TERESA
TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR TRANSPOSITION USING
TITLE OF INVENTION: MINIMAL SEGMENTS OF THE EUKARYOTIC TRANSFORMATION
FILE REFERENCE: 835910-92098
CURRENT APPLICATION NUMBER: US/10/001,189
PRIOR FILING DATE: 2001-10-30
PRIOR APPLICATION NUMBER: 60/244,984
PRIOR FILING DATE: 2000-11-01
PRIOR APPLICATION NUMBER: 60/244,677
PRIOR FILING DATE: 2000-10-31
NUMBER OF SEQ ID NOS: 70
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 45
LENGTH: 6984
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Description of Artificial Sequence:
OTHER INFORMATION: pBXP3-Dsred-ori sequence
US-10-001-189-45

Query Match 60.1%; Score 142.4; DB 9; Length 6984;
Best Local Similarity 99.3%; Pred. No. 1.6e-34;
Matches 143; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 94 TGCATCTCAATTAGTCAGCAACCATAGTCCCGCCCTAACTCCGCCCATCCGCCCTAA 153
DB 4649 TGCATCTCAATTAGTCAGCAACCATAGTCCCGCCCTAACTCCGCCCATCCGCCCTAA 4708
QY 154 CTCGCCCGAGTTCGCCCGCATTCCTCGGCCCATCGCTGACTAATTTTTTTTATTCGAG 213
DB 4709 CTCGCCCGAGTTCGCCCGCATTCCTCGGCCCATCGCTGACTAATTTTTTTTATTCGAG 4768
QY 214 AGCGCGAGCGCCCTCGGCCCTCTG 237
DB 4769 AGCGCGAGCGCCCTCGGCCCTCTG 4792

RESULT 13
US-08-779-457-46
; Sequence 46, Application US/08779457
; Publication No. US20020193571A1
; GENERAL INFORMATION:
; APPLICANT: Carter, Paul J.
; APPLICANT: Chiang, Nancy Y.
; APPLICANT: Kyung, Jin Kim
; APPLICANT: Matthews, William
; APPLICANT: Rodrigues, Maria L.
TITLE OF INVENTION: WSX RECEPTOR AGONIST ANTIBODIES

us-09-787-562-9.rnpb

Mon Jan 6 15:20:25 2003

NUMBER OF SEQUENCES: 51
CORRESPONDENCE ADDRESS:
ADDRESSEE: Genentech, Inc.
STREET: 460 Point San Bruno Blvd
CITY: South San Francisco
STATE: California
COUNTRY: USA
ZIP: 94080
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: WinPatIn (Genentech)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/779,457
FILING DATE: 01/08/96
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/667197
FILING DATE: 06/20/96
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/585005
FILING DATE: 01/08/96
ATTORNEY/AGENT INFORMATION:
NAME: Lee, Wendy M.
REGISTRATION NUMBER: 40,378
REFERENCE/DOCKET NUMBER: P0986P2
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415/225-1994
TELEFAX: 415/952-9881
TELEX: 910/371-7168
INFORMATION FOR SEQ ID NO: 46:
SEQUENCE CHARACTERISTICS:
LENGTH: 7127 base pairs
TYPE: Nucleic Acid
STRANDEDNESS: Double
TOPOLOGY: Linear
US-08-779-457-46

Query Match 60.1%; Score 142.4; DB 7; Length 7127;
Best Local Similarity 99.3%; Pred. No. 1.6e-34;
Matches 143; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 94 TGCATCTCAATTAGTCAGCACCATAGTCGCCGCCCTAACTCCGCCCATCCGCCCTAA 153
DB 2551 TGCATCTCAATTAGTCAGCACCATAGTCGCCGCCCTAACTCCGCCCATCCGCCCTAA 2610
QY 154 CTCGCCCCAGTTCGCCCCCATCTCCGCCCATCGCTGACTAAATTTTATTTATGCGAG 213
DB 2611 CTCGCCCCAGTTCGCCCCCATCTCCGCCCATCGCTGACTAAATTTTATTTATGCGAG 2670
QY 214 AGCGCGAGCGCGCTCGGCCCTCTG 237
DB 2671 AGCGCGAGCGCGCTCGGCCCTCTG 2694

RESULT 14
US-09-982-610-19
Sequence 19, Application US/09982610
Patent No. US20020146420A1
GENERAL INFORMATION:
APPLICANT: Genentech, Inc.
Bennett, Brian D.
Goeddel, David
Lee, James M.
Matthews, William
Tsai, Siao Ping
Wood, William I.
TITLE OF INVENTION: PROTEIN TYROSINE KINASE AGONIST ANTIBODIES
NUMBER OF SEQUENCES: 45
CORRESPONDENCE ADDRESS:
ADDRESSEE: Genentech, Inc.
STREET: 460 Point San Bruno Blvd

CITY: South San Francisco
STATE: California
COUNTRY: USA
ZIP: 94080
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: WinPatIn (Genentech)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/982,610
FILING DATE: 17-Oct-2001
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/446,648
FILING DATE: 1996-MAY-23
APPLICATION NUMBER: 08/222616
FILING DATE: 04-APR-1994
ATTORNEY/AGENT INFORMATION:
NAME: Lee, Wendy M.
REGISTRATION NUMBER: 40,378
REFERENCE/DOCKET NUMBER: P0821P3PCT
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415/225-1994
TELEFAX: 415/952-9881
TELEX: 910/371-7168
INFORMATION FOR SEQ ID NO: 19:
SEQUENCE CHARACTERISTICS:
LENGTH: 7607 base pairs
TYPE: Nucleic Acid
STRANDEDNESS: Single
TOPOLOGY: Linear
SEQUENCE DESCRIPTION: SEQ ID NO: 19:
US-09-982-610-19

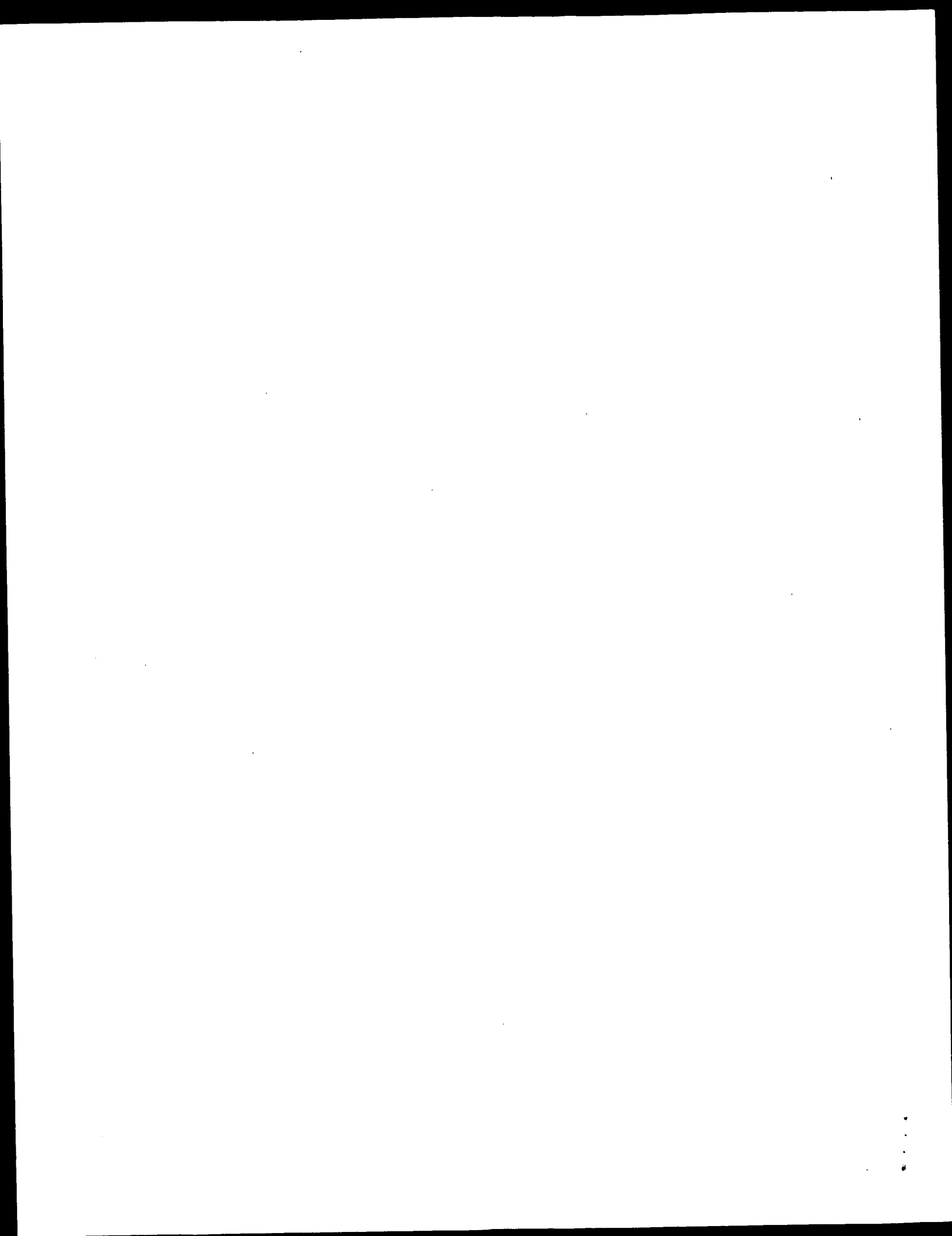
Query Match 60.1%; Score 142.4; DB 10; Length 7607;
Best Local Similarity 99.3%; Pred. No. 1.7e-34;
Matches 143; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 94 TGCATCTCAATTAGTCAGCACCATAGTCGCCGCCCTAACTCCGCCCATCCGCCCTAA 153
DB 4284 TGCATCTCAATTAGTCAGCACCATAGTCGCCGCCCTAACTCCGCCCATCCGCCCTAA 4343
QY 154 CTCGCCCCAGTTCGCCCCCATCTCCGCCCATCGCTGACTAAATTTTATTTATGCGAG 213
DB 4344 CTCGCCCCAGTTCGCCCCCATCTCCGCCCATCGCTGACTAAATTTTATTTATGCGAG 4403
QY 214 AGCGCGAGCGCGCTCGGCCCTCTG 237
DB 4404 AGCGCGAGCGCGCTCGGCCCTCTG 4427

RESULT 15
US-09-828-825-1
Sequence 1, Application US/09828825
Patent No. US20020018767A1
GENERAL INFORMATION:
APPLICANT: Lee, Seewoo
APPLICANT: Kim, Han-soo
TITLE OF INVENTION: Anti-cancer Cellular Vaccine
FILE REFERENCE: 84906-102
CURRENT APPLICATION NUMBER: US/09/828,825
CURRENT FILING DATE: 2001-04-10
PRIOR APPLICATION NUMBER: KR00-43498
PRIOR FILING DATE: 2000-07-27
NUMBER OF SEQ ID NOS: 16
SOFTWARE: Patent In Ver. 2.1
SEQ ID NO 1
LENGTH: 8578
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
NAME/KEY: enhancer

LOCATION: (1)..(659)
OTHER INFORMATION: CMV IE
NAME/KEY: promoter
LOCATION: (669)..(750)
OTHER INFORMATION: CMV IE
NAME/KEY: promoter
LOCATION: (1067)..(1085)
OTHER INFORMATION: T7 RNA Promoter
NAME/KEY: mRNA
LOCATION: (1090)..(1984)
OTHER INFORMATION: Human B7.1 (1090-1956 is coding sequence)
NAME/KEY: RBS
LOCATION: (2013)..(2593)
OTHER INFORMATION: IRES sequence
NAME/KEY: mRNA
LOCATION: (2627)..(4263)
OTHER INFORMATION: Human IL12.0 (2640-4223 coding sequence, flexible
OTHER INFORMATION: linker at 3624-3629)
NAME/KEY: promoter
LOCATION: (4352)..(4431)
OTHER INFORMATION: T3 RNA polymerase promoter
NAME/KEY: polyA_signal
LOCATION: (4362)..(4583)
OTHER INFORMATION: SV40 fragment containing polyadenylation signal
NAME/KEY: rep_origin
LOCATION: (4678)..(5133)
OTHER INFORMATION: fl origin of replication
NAME/KEY: misc_feature
LOCATION: (5197)..(6564)
OTHER INFORMATION: Neo r expression cassette
NAME/KEY: misc_feature
LOCATION: (6975)..(7835)
OTHER INFORMATION: Ampicillin resistance
OTHER INFORMATION: Description of Artificial Sequence: Plasmid
NAME/KEY: CDS
LOCATION: (2640)..(4223)
OTHER INFORMATION: IL 12.0 coding sequence - 2 amino acid linker at
OTHER INFORMATION: 3624
US-09-828-825-1

Query Match 60.1%; Score 142.4; DB 10; Length 8578;
Best Local Similarity 99.3%; Pred. No. 1.7e-34;
Matches 143; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 94 TGCATCTCAATTAGTCAGCAACCATAGTCCCGCCCTAACTCCGCCCATCCGCCCTAA 153
|||||
Db 5322 TGCATCTCAATTAGTCAGCAACCATAGTCCCGCCCTAACTCCGCCCATCCGCCCTAA 5381
|||||
QY 154 CTCGCCCGAGTTCCGCCCATCTCCGCCCATCGCTGACTAATTTTATTTATGCAG 213
|||||
Db 5382 CTCGCCCGAGTTCCGCCCATCTCCGCCCATCGCTGACTAATTTTATTTATGCAG 5441
|||||
QY 214 AGCGCGAGCGCGCTCGGCTCTG 237
|||||
Db 5442 AGCGCGAGCGCGCTCGGCTCTG 5465

Search completed: January 4, 2003, 01:06:16
Job time : 88.9968 secs



GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 3, 2003, 22:52:46 ; Search time 108.076 Seconds
(without alignments)
3231.380 Million cell updates/sec

Title: US-09-787-562-10
Perfect score: 12
Sequence: 1 gtcgtcaggca 12

Scoring table: IDENTITY_NUC
Gapop 10.0, Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

GenEmbl.*

1: gb_ba.*
2: gb_htg.*
3: gb_in.*
4: gb_om.*
5: gb_ov.*
6: gb_pat.*
7: gb_ph.*
8: gb_pi.*
9: gb_pr.*
10: gb_ro.*
11: gb_sts.*
12: gb_sy.*
13: gb_un.*
14: gb_vi.*
15: em_ba.*
16: em_fun.*
17: em_hum.*
18: em_in.*
19: em_mu.*
20: em_om.*
21: em_or.*
22: em_ov.*
23: em_pat.*
24: em_ph.*
25: em_pl.*
26: em_ro.*
27: em_sts.*
28: em_un.*
29: em_vi.*
30: em_htg_hum.*
31: em_htg_inv.*
32: em_htg_other.*
33: em_htg_mus.*
34: em_htg_pln.*
35: em_htg_rod.*
36: em_htg_mam.*
37: em_htg_vrt.*
38: em_sy.*
39: em_htgo_hum.*
40: em_htgo_mus.*
41: em_htgo_other.*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Match	Query Length	DB ID	Description
1	12	100.0	12	6	AX023668 Sequence
2	12	100.0	24	6	AX023669 Sequence
3	12	100.0	35	6	I58420 Sequence 4
4	12	100.0	36	6	I58427 Sequence 11
5	12	100.0	170	6	E01480 Sequence 1
6	12	100.0	170	6	E01480 DNA encodin
7	12	100.0	174	6	E01981 DNA encodin
8	12	100.0	177	6	I58418 Sequence 1
9	12	100.0	177	6	I09058 Sequence 36
10	12	100.0	183	6	A00372 Artificial
11	12	100.0	183	6	A00373 Artificial
12	12	100.0	185	6	E01479 DNA sequenc
13	12	100.0	185	6	E01980 DNA encodin
14	12	100.0	185	6	E02112 DNA sequenc
15	12	100.0	191	6	E02089 DNA encodin
16	12	100.0	191	6	E02091 DNA encodin
17	12	100.0	237	6	AX023667 Sequence
18	12	100.0	269	11	PAU15405
19	12	100.0	296	10	MMU31551
20	12	100.0	338	6	E02090
21	12	100.0	509	1	AF528038
22	12	100.0	515	6	E02113
23	12	100.0	575	6	AX01316 Sequence
24	12	100.0	617	9	HS4326857
25	12	100.0	825	8	AF108439
26	12	100.0	825	8	AY086953
27	12	100.0	987	3	DME238007
28	12	100.0	1253	1	AF005670
29	12	100.0	1256	1	AF005673
30	12	100.0	1256	1	AF005674
31	12	100.0	1625	1	MTV001
32	12	100.0	1818	9	BC005158
33	12	100.0	1845	6	AX123571
34	12	100.0	1968	6	AX066951
35	12	100.0	1968	6	AX066953
36	12	100.0	1985	9	HS805560
37	12	100.0	2013	9	AK090848
38	12	100.0	2064	9	AK095929
39	12	100.0	2087	6	ARI46835
40	12	100.0	2103	9	AF189270
41	12	100.0	2219	1	XORPFCGEN
42	12	100.0	2360	8	GPU93077
43	12	100.0	2564	9	PATLR3
44	12	100.0	2592	8	AF168786
45	12	100.0	2624	10	BC020313
			2713	1	REU278372

ALIGNMENTS

RESULT 1
AX023668
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL

AX023668
Sequence 10 from Patent WO0017371.
AX023668
AX023668.1 GI:10184029

synthetic construct.
synthetic construct
artificial sequences.
1 (bases 1 to 12)

Binley,K.M. and Naylor,S.
Polynucleotide constructs and uses thereof
Patent: WO 0017371-A 10 30-MAR-2000;
BINLEY KATIE MARY (GB) ; NAYLOR STUART (GB) ; OXFORD BIOMEDICA LTD

```

FEATURES             (GB)             Location/Qualifiers
source               1..12
/organism="synthetic construct"
/db_xref="taxon:32630"
/note="Spacer"
BASE COUNT          2 a   3 c   5 g   2 t
ORIGIN
Query Match          100.0%; Score 12; DB 6; Length 12;
Best Local Similarity 100.0%; Pred. No. 1.1e+04;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GTCGTGCAGGCA 12
|||||
Db 1 GTCGTGCAGGCA 13

RESULT 4
I58427
LOCUS               36 bp      DNA      linear      PAT 07-OCT-1997
DEFINITION          Sequence 11 from patent US 5652120.
ACCESSION            I58427
VERSION              I58427.1 GI:2477665
KEYWORDS             Unknown.
SOURCE               Unknown.
ORGANISM             Unclassified.
REFERENCE            1 (bases 1 to 36)
AUTHORS              Park,S.Kook., Lee,K.Moon., Nho,K.Seung., Koh,Y.Wook., Kwon,C.Hyuk.,
                    Chung,J.Young., Jee,Y.Su. and Yu,Y.Hyo.
TITLE                Gene coding human epidermal growth factor and process for preparing
                    the same
JOURNAL              Patent: US 5652120-A 11 29-JUL-1997;
FEATURES             Location/Qualifiers
source               1..36
                    /organism="unknown"
BASE COUNT          8 a   9 c   11 g   8 t
ORIGIN
Query Match          100.0%; Score 12; DB 6; Length 36;
Best Local Similarity 100.0%; Pred. No. 1e+04;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GTCGTGCAGGCA 12
|||||
Db 21 GTCGTGCAGGCA 32

RESULT 5
E01480/c
LOCUS               170 bp      DNA      linear      PAT 29-SEP-1997
DEFINITION          DNA sequence encoding EGF derivative in recombinant cloning vector.
ACCESSION            E01480
VERSION              E01480.1 GI:2169736
KEYWORDS             JP 1988003791-A/1.
SOURCE               Synthetic construct.
ORGANISM             artificial sequences.
REFERENCE            1 (bases 1 to 170)
AUTHORS              Matsui,A., Harada,Y. and Nakano,T.
TITLE                RECOMBINANT PLASMID
JOURNAL              Patent: JP 1988003791-A 1 08-JAN-1988;
                    HITACHI LTD, HITACHI CHEM CO LTD
COMMENT              OS Artificial gene
                    PN JP 1988003791-A/1
                    PD 08-JAN-1988
                    PF 25-JUN-1986 JP 1986146964
                    PI MATSUI AKIKO HARADA YOSHINORI, NAKANO TAKAMORI PC
                    C12N15/00/C07H21/04,C12P21/02,C12P21/02,C12R1.19; CC
                    CC topology: Linear;
                    CC hypothetical: No;
                    CC anti-sense: No;
                    Key Location/Qualifiers
                    FH CDS      18..170
                    FH FT       /product='epidermal growth factor derivative'
                    FT RBS      6..9
                    FT misc_signal 11..16
                    FT FT       /note='XbaI recognition site'.

FEATURES             Location/Qualifiers
source               1..24
/organism="synthetic construct"
/db_xref="taxon:32630"
/note="Spacer"
BASE COUNT          4 a   5 c   7 g   8 t
ORIGIN
Query Match          100.0%; Score 12; DB 6; Length 24;
Best Local Similarity 100.0%; Pred. No. 1e+04;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GTCGTGCAGGCA 12
|||||
Db 7 GTCGTGCAGGCA 18

RESULT 3
I58420/c
LOCUS               35 bp      DNA      linear      PAT 07-OCT-1997
DEFINITION          Sequence 4 from patent US 5652120.
ACCESSION            I58420
VERSION              I58420.1 GI:2477658
KEYWORDS             Unknown.
SOURCE               Unknown.
ORGANISM             Unclassified.
REFERENCE            1 (bases 1 to 35)
AUTHORS              Park,S.Kook., Lee,K.Moon., Nho,K.Seung., Koh,Y.Wook., Kwon,C.Hyuk.,
                    Chung,J.Young., Jee,Y.Su. and Yu,Y.Hyo.
TITLE                Gene coding human epidermal growth factor and process for preparing
                    the same
JOURNAL              Patent: US 5652120-A 4 29-JUL-1997;
FEATURES             Location/Qualifiers
source               1..35
                    /organism="unknown"
BASE COUNT          7 a   11 c   10 g   7 t
ORIGIN

```

```

FEATURES
  source
    Location/Qualifiers
    1..170
    /organism="synthetic construct"
    /db_xref="taxon:32630"
  BASE COUNT      44 a      41 c      45 g      40 t
  ORIGIN
    Query Match      100.0%; Score 12; DB 6; Length 170;
    Best Local Similarity 100.0%; Pred. No. 9.4e+03;
    Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

  QY 1 GTCGTGCAGGCA 12
  Db 71 GTCGTGCAGGCA 60

  RESULT 6
  E01981/c
  LOCUS      E01981      170 bp      RNA      linear      PAT 29-SEP-1997
  DEFINITION DNA encoding epidermal growth factor.
  ACCESSION E01981
  VERSION E01981.1 GI:2170229
  KEYWORDS JP 1989132383-A/2.
  SOURCE Unidentified.
  ORGANISM Homo sapiens
    Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
    Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
  REFERENCE
  AUTHORS Matsui,A. and Harada,Y.
  TITLE PLASMID RECOMBINANT
  JOURNAL Patent: JP 1989132383-A 2 24-MAY-1989;
  COMMENT HITACHI LTD HITACHI CHEM CO LTD
  PN JP 1989132383-A/2
  PD 24-MAY-1989
  PF 18-NOV-1987 JP 1987289332
  PI MATSUI AKIKO, HARADA YOSHINORI
  PC C12N15/00,(C12N15/00,C12R1:91);
  CC strandedness: Double;
  CC topology: Linear;
  CC hypothetical: No;
  CC anti-sense: No;
  CC *source: clone=pBR322 UG;
  CC Feature is identified by experimental;
  FH Key Location/Qualifiers
  PH
  FT 5'UTR 1..17
  FT CDS 18..167
  FT /product='epidermal growth factor' FT
  mat_peptide 18..164
  FT /product='epidermal growth factor'.
  FEATURES
  source
    Location/Qualifiers
    1..170
    /organism="Homo sapiens"
    /db_xref="taxon:9606"
  BASE COUNT      44 a      41 c      45 g      40 t
  ORIGIN
    Query Match      100.0%; Score 12; DB 6; Length 170;
    Best Local Similarity 100.0%; Pred. No. 9.4e+03;
    Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

  QY 1 GTCGTGCAGGCA 12
  Db 71 GTCGTGCAGGCA 60

  RESULT 7
  I58418/c
  LOCUS      I58418      174 bp      DNA      linear      PAT 07-OCT-1997
  DEFINITION Sequence 1 from patent US 5652120.
  ACCESSION I58418
  VERSION I58418.1 GI:2477656
  KEYWORDS

  SOURCE Unknown.
  ORGANISM Unclassified.
  REFERENCE 1 (bases 1 to 174)
  AUTHORS Park,S.Kook., Lee,K.Moon., Nho,K.Seung., Koh,Y.Wook., Kwon,C.Hyuk.,
  Chung,J.Young., Jee,Y.Su. and Yu,Y.Hyo.
  TITLE Gene coding human epidermal growth factor and process for preparing
  the same
  JOURNAL Patent: US 5652120-A 1 29-JUL-1997;
  FEATURES
  Location/Qualifiers
  1..174
  /organism="unknown"
  BASE COUNT      40 a      47 c      50 g      37 t
  ORIGIN
    Query Match      100.0%; Score 12; DB 6; Length 174;
    Best Local Similarity 100.0%; Pred. No. 9.4e+03;
    Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

  QY 1 GTCGTGCAGGCA 12
  Db 54 GTCGTGCAGGCA 43

  RESULT 8
  I09058/c
  LOCUS      I09058      177 bp      DNA      linear      PAT 02-DEC-1994
  DEFINITION Sequence 36 from Patent WO 8809344.
  ACCESSION I09058
  VERSION I09058.1 GI:588241
  KEYWORDS
  SOURCE Unknown.
  ORGANISM Unclassified.
  REFERENCE 1 (bases 1 to 177)
  AUTHORS Huston,J.S. and Oppermann,H.
  TITLE TARGETED MULTIFUNCTIONAL PROTEINS
  JOURNAL Patent: WO 8809344-A 36 01-DEC-1988;
  FEATURES
  Location/Qualifiers
  1..177
  /organism="unknown"
  BASE COUNT      39 a      45 c      53 g      40 t
  ORIGIN
    Query Match      100.0%; Score 12; DB 6; Length 177;
    Best Local Similarity 100.0%; Pred. No. 9.4e+03;
    Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

  QY 1 GTCGTGCAGGCA 12
  Db 60 GTCGTGCAGGCA 49

  RESULT 9
  A00372/c
  LOCUS      A00372      183 bp      DNA      linear      PAT 04-MAR-1993
  DEFINITION Artificial gene for human epidermal growth factor.
  ACCESSION A00372
  VERSION A00372.1 GI:344185
  KEYWORDS epidermal growth factor.
  SOURCE synthetic construct.
  ORGANISM artificial construct.
  FEATURES
  Location/Qualifiers
  1..183
  /organism="synthetic construct"
  /db_xref="taxon:32630"
  10..174
  /codon_start=1
  /transl_table=11
  /product="epidermal growth factor"
  /protein_id="CAA00059.1"
  /db_xref="GI:344186"

```

```

/translation="MNSDECPUSHDGYCLHDGVCVMYIHALDKYACNCVVGIGERCQ
YRLKWWELR"
BASE COUNT      47 a      45 c      42 t
ORIGIN
Query Match      100.0%; Score 12; DB 6; Length 183;
Best Local Similarity 100.0%; Pred. No. 9.4e+03;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GTCGTGCAGGCA 12
    |||
Db 63 GTCGTGCAGGCA 52

RESULT 10
A00373      A00373      183 bp      DNA      linear      PAT 04-MAR-1993
LOCUS      Artificial gene for human epidermal growth factor, reverse
DEFINITION
ACCESSION      A00373
VERSION      A00373.1 GI:344187
KEYWORDS      synthetic construct.
SOURCE      synthetic construct
ORGANISM      artificial sequences.
FEATURES
    source      Location/Qualifiers
        1..183
        /db_xref="taxon:32630"
        /organism="synthetic construct"
        /db_xref="taxon:32630"
BASE COUNT      42 a      49 c      45 g      47 t
ORIGIN
Query Match      100.0%; Score 12; DB 6; Length 183;
Best Local Similarity 100.0%; Pred. No. 9.4e+03;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GTCGTGCAGGCA 12
    |||
Db 121 GTCGTGCAGGCA 132

RESULT 11
E01479/c
LOCUS      E01479      185 bp      DNA      linear      PAT 29-SEP-1997
DEFINITION      DNA sequence encoding EGF in recombinant cloning vector.
ACCESSION      E01479
VERSION      E01479.1 GI:2169735
KEYWORDS      JP 1988003790-A/1.
SOURCE      synthetic construct.
ORGANISM      artificial sequences.
REFERENCE      1 (bases 1 to 185)
AUTHORS      Harada, Y., Matsui, A. and Ito, M.
TITLE      RECOMBINANT PLASMID
JOURNAL      Patent: JP 1988003790-A 1 08-JAN-1988;
            HITACHI LTD, HITACHI CHEM CO LTD
COMMENT      OS Artificial gene
            OC Artificial sequence; Genes.
            PN JP 1988003790-A/1
            PD 08-JAN-1988
            PE 25-JUN-1986 JP 1986146963
            PI HARADA YOSHINORI, MATSUI AKIKO, ITO MICHIO
            PC C12N15/00/C07H21/04,C12P21/02,C12R1:19; CC
            strandedness: Double;
            CC topology: Linear;
            CC hypothetical: No;
            CC anti-sense: No;
            FH Key      Location/Qualifiers
            FH CDS      18..185
            FT          /product='epidermal growth factor' FT
            FT          misc_signal 11..16

/translation="MNSDECPUSHDGYCLHDGVCVMYIHALDKYACNCVVGIGERCQ
YRLKWWELR"
BASE COUNT      47 a      45 c      42 t
ORIGIN
Query Match      100.0%; Score 12; DB 6; Length 183;
Best Local Similarity 100.0%; Pred. No. 9.4e+03;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GTCGTGCAGGCA 12
    |||
Db 63 GTCGTGCAGGCA 52

RESULT 10
A00373      A00373      183 bp      DNA      linear      PAT 04-MAR-1993
LOCUS      Artificial gene for human epidermal growth factor, reverse
DEFINITION
ACCESSION      A00373
VERSION      A00373.1 GI:344187
KEYWORDS      synthetic construct.
SOURCE      synthetic construct
ORGANISM      artificial sequences.
FEATURES
    source      Location/Qualifiers
        1..183
        /db_xref="taxon:32630"
        /organism="synthetic construct"
        /db_xref="taxon:32630"
BASE COUNT      42 a      49 c      45 g      47 t
ORIGIN
Query Match      100.0%; Score 12; DB 6; Length 183;
Best Local Similarity 100.0%; Pred. No. 9.4e+03;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GTCGTGCAGGCA 12
    |||
Db 121 GTCGTGCAGGCA 132

RESULT 11
E01479/c
LOCUS      E01479      185 bp      DNA      linear      PAT 29-SEP-1997
DEFINITION      DNA sequence encoding EGF in recombinant cloning vector.
ACCESSION      E01479
VERSION      E01479.1 GI:2169735
KEYWORDS      JP 1988003790-A/1.
SOURCE      synthetic construct.
ORGANISM      artificial sequences.
REFERENCE      1 (bases 1 to 185)
AUTHORS      Harada, Y., Matsui, A. and Ito, M.
TITLE      RECOMBINANT PLASMID
JOURNAL      Patent: JP 1988003790-A 1 08-JAN-1988;
            HITACHI LTD, HITACHI CHEM CO LTD
COMMENT      OS Artificial gene
            OC Artificial sequence; Genes.
            PN JP 1988003790-A/1
            PD 08-JAN-1988
            PE 25-JUN-1986 JP 1986146963
            PI HARADA YOSHINORI, MATSUI AKIKO, ITO MICHIO
            PC C12N15/00/C07H21/04,C12P21/02,C12R1:19; CC
            strandedness: Double;
            CC topology: Linear;
            CC hypothetical: No;
            CC anti-sense: No;
            FH Key      Location/Qualifiers
            FH CDS      18..185
            FT          /product='epidermal growth factor' FT
            FT          misc_signal 11..16

```

```

FT          /note='XbaI recognition site'.
FEATURES
    source      Location/Qualifiers
        1..185
        /organism="synthetic construct"
        /db_xref="taxon:32630"
BASE COUNT      46 a      44 c      52 g      43 t
ORIGIN
Query Match      100.0%; Score 12; DB 6; Length 185;
Best Local Similarity 100.0%; Pred. No. 9.3e+03;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GTCGTGCAGGCA 12
    |||
Db 71 GTCGTGCAGGCA 60

RESULT 12
E01980/c
LOCUS      E01980      185 bp      RNA      linear      PAT 29-SEP-1997
DEFINITION      DNA encoding epidermal growth factor.
ACCESSION      E01980
VERSION      E01980.1 GI:2170228
KEYWORDS      JP 1989132383-A/1.
SOURCE      Unidentified.
ORGANISM      Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE      1 (bases 1 to 185)
AUTHORS      Matsui, A. and Harada, Y.
TITLE      PLASMID RECOMBINANT
JOURNAL      Patent: JP 1989132383-A 1 24-MAY-1989;
            HITACHI LTD, HITACHI CHEM CO LTD
COMMENT      PN JP 1989132383-A/1
            PD 24-MAY-1989
            PF 18-NOV-1987 JP 1987289332
            PI MATSUI AKIKO, HARADA YOSHINORI
            PC C12N15/00,C12N15/00,C12R1:91;
            CC strandedness: Double;
            CC topology: Linear;
            CC hypothetical: No;
            CC anti-sense: No;
            CC *source: clone-pBR 322 UG;
            CC feature is identified by experimental;
            FH Key      Location/Qualifiers
            FH 5'UTR      1..17
            FH CDS      18..182
            FT          /product='epidermal growth factor' FT
            FT          mat_peptide 18..179
            FT          /product='epidermal growth factor'.
FEATURES
    source      Location/Qualifiers
        1..185
        /organism="Homo sapiens"
        /db_xref="taxon:9606"
BASE COUNT      46 a      44 c      52 g      43 t
ORIGIN
Query Match      100.0%; Score 12; DB 6; Length 185;
Best Local Similarity 100.0%; Pred. No. 9.3e+03;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GTCGTGCAGGCA 12
    |||
Db 71 GTCGTGCAGGCA 60

RESULT 13
E02112/c
LOCUS      E02112      185 bp      DNA      linear      PAT 29-SEP-1997
DEFINITION      DNA sequence coding for human epidermal growth factor(hEGF).
ACCESSION      E02112
VERSION      E02112.1 GI:2170354

```

```

KEYWORDS      JP 1989257481-A/1.
SOURCE        Homo sapiens.
ORGANISM      Mammalia; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
              Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE     1. (bases 1 to 185)
AUTHORS       Harada,Y., and Matsui,A.
TITLE         DNA HAVING SYNTHETIC GENE FOR PRODUCING HUMAN EPIDERMAL GROWTH
              FACTOR AND PLASMID RECOMBINANT THEREOF
JOURNAL       Patent: JP 1989257481-A 1 13-OCT-1989;
              HITACHI LTD, HITACHI CHEM CO LTD
COMMENT       OS Homo sapiens (Human)
              PN JP 1989257481-A/1
              PD 13-OCT-1989
              PF 08-APR-1988 JP 1988085072
              PI HARADA YOSHINORI, MATSUI AKIKO
              PC C12N15/00;
              CC strandedness: Double;
              CC topology: Linear;
              CC hypothetical: No;
              CC anti-sense: No;
              FH Key Location/Qualifiers
              FT 5'UTR 1..17
              FT CDS 18..185
              FT misc_signal 6..9
              FT Location/Qualifiers
FEATURES      1..185
              source      Location/Qualifiers
              1..185
              /organism="Homo sapiens"
              /db_xref="taxon:9606"
BASE COUNT   45 a 43 c 54 g 43 t
ORIGIN
Query Match      100.0%; Score 12; DB 6; Length 185;
Best Local Similarity 100.0%; Pred. No. 9.3e+03;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GTCGTGCAGGCA 12
|||||
Db 71 GTCGTGCAGGCA 60

RESULT 14
E02089/c
LOCUS
DEFINITION    DNA encoding human epidermal growth factor (hEGF).
ACCESSION     E02089
VERSION       E02089.1 GI:2170331
KEYWORDS      JP 1989247098-A/1.
SOURCE        Synthetic construct.
ORGANISM      artificial sequences.
REFERENCE     1 (bases 1 to 191)
AUTHORS       Harada,Y., Shimizu,N., Fukuzono,S. and Fujimori,K.
TITLE         BIOTECHNOLOGICAL PRODUCTION OF HUMAN EPITHELIOCYTE GROWTH FACTOR
JOURNAL       Patent: JP 1989247098-A 1 02-OCT-1989;
              HITACHI LTD, HITACHI CHEM CO LTD
COMMENT       OS Artificial gene
              OC Artificial sequence; Genes.
              OS Human
              PN JP 1989247098-A/1
              PD 02-OCT-1989
              PF 30-MAR-1988 JP 1988074382
              PI HARADA YOSHINORI, SHIMIZU NORIO, FUKUZONO SHINICHI, PI
              FUJIMORI KIYOSHI
              PC C12P21/02,C07H21/04,C12N15/00,(C12P21/02,C12R1:19); CC
              strandedness: Double;
              CC topology: Linear;
              CC hypothetical: No;
              CC anti-sense: No;
              CC *source: tissue_type=duodenum;
              CC *source: clone=pbREG;
              CC Feature is identified by similarity;
              FH Key Location/Qualifiers
              FT CDS 24..191
              FT /product='human epidermal growth factor' FT
              FT mat_peptide 27..185.
              FT Location/Qualifiers
              1..191
              /organism="synthetic construct"
              /db_xref="taxon:32630"
BASE COUNT   48 a 45 c 53 g 45 t
ORIGIN
Query Match      100.0%; Score 12; DB 6; Length 191;
Best Local Similarity 100.0%; Pred. No. 9.3e+03;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GTCGTGCAGGCA 12
|||||
Db 77 GTCGTGCAGGCA 66

RESULT 15
E02091/c
LOCUS
DEFINITION    DNA encoding human epidermal growth factor (hEGF).
ACCESSION     E02091
VERSION       E02091.1 GI:2170333
KEYWORDS      JP 1989247099-A/1.
SOURCE        Synthetic construct.
ORGANISM      artificial sequences.
REFERENCE     1 (bases 1 to 191)
AUTHORS       Shimizu,N., Harada,Y., Fukuzono,S. and Fujimori,K.
TITLE         BIOTECHNOLOGICAL PRODUCTION OF HUMAN EPITHELIOCYTE GROWTH FACTOR
JOURNAL       Patent: JP 1989247099-A 1 02-OCT-1989;
              HITACHI LTD, HITACHI CHEM CO LTD
COMMENT       OS Artificial gene
              OC Artificial sequence; Genes.
              OS Human
              PN JP 1989247099-A/1
              PD 02-OCT-1989
              PF 30-MAR-1988 JP 1988074383
              PI SHIMIZU NORIO, HARADA YOSHINORI, FUKUZONO SHINICHI, PI
              FUJIMORI KIYOSHI
              PC C12P21/02,C07H21/04,C12N15/00,(C12P21/02,C12R1:19); CC
              strandedness: Double;
              CC topology: Linear;
              CC hypothetical: No;
              CC anti-sense: No;
              CC *source: tissue_type=duodenum;
              CC *source: clone=pbREG;
              CC Feature is identified by similarity;
              FH Key Location/Qualifiers
              FT CDS 24..191
              FT /product='human epidermal growth factor' FT
              FT mat_peptide 27..185.
              FT Location/Qualifiers
              1..191
              /organism="synthetic construct"
              /db_xref="taxon:32630"
BASE COUNT   48 a 45 c 53 g 45 t
ORIGIN
Query Match      100.0%; Score 12; DB 6; Length 191;
Best Local Similarity 100.0%; Pred. No. 9.3e+03;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GTCGTGCAGGCA 12
|||||
Db 77 GTCGTGCAGGCA 66

RESULT 15
E02091/c
LOCUS
DEFINITION    DNA encoding human epidermal growth factor (hEGF).
ACCESSION     E02091
VERSION       E02091.1 GI:2170333
KEYWORDS      JP 1989247099-A/1.
SOURCE        Synthetic construct.
ORGANISM      artificial sequences.
REFERENCE     1 (bases 1 to 191)
AUTHORS       Shimizu,N., Harada,Y., Fukuzono,S. and Fujimori,K.
TITLE         BIOTECHNOLOGICAL PRODUCTION OF HUMAN EPITHELIOCYTE GROWTH FACTOR
JOURNAL       Patent: JP 1989247099-A 1 02-OCT-1989;
              HITACHI LTD, HITACHI CHEM CO LTD
COMMENT       OS Artificial gene
              OC Artificial sequence; Genes.
              OS Human
              PN JP 1989247099-A/1
              PD 02-OCT-1989
              PF 30-MAR-1988 JP 1988074383
              PI SHIMIZU NORIO, HARADA YOSHINORI, FUKUZONO SHINICHI, PI
              FUJIMORI KIYOSHI
              PC C12P21/02,C07H21/04,C12N15/00,(C12P21/02,C12R1:19); CC
              strandedness: Double;
              CC topology: Linear;
              CC hypothetical: No;
              CC anti-sense: No;
              CC *source: tissue_type=duodenum;
              CC *source: clone=pbREG;
              CC Feature is identified by similarity;
              FH Key Location/Qualifiers
              FT CDS 24..191
              FT /product='human epidermal growth factor' FT
              FT mat_peptide 27..185.
              FT Location/Qualifiers
              1..191
              /organism="synthetic construct"
              /db_xref="taxon:32630"
BASE COUNT   48 a 45 c 53 g 45 t
ORIGIN
Query Match      100.0%; Score 12; DB 6; Length 191;
Best Local Similarity 100.0%; Pred. No. 9.3e+03;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GTCGTGCAGGCA 12
|||||
Db 77 GTCGTGCAGGCA 66

```

Mon Jan 6 15:20:18 2003

Page 6

us-09-787-562-10.rge

Db 77 GTCTGCGAGCA 66
|||||

Search completed: January 3, 2003, 23:54:42
Job time : 111.076 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 3, 2003, 22:52:02 ; Search time 20.0631 Seconds
(without alignments)
1346.950 Million cell updates/sec

Title: US-09-787-562-10

Perfect score: 12

Sequence: 1 gtcgtgagca 12

Scoring table: IDENTITY_NUC

Gapop 10.0, Gapext 1.0

Searched: 2185239 seqs, 1125999159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- N_Geneseq_101002.*
- 1: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1980.DAT.*
 - 2: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1981.DAT.*
 - 3: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1982.DAT.*
 - 4: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1983.DAT.*
 - 5: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1984.DAT.*
 - 6: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1985.DAT.*
 - 7: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1986.DAT.*
 - 8: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1987.DAT.*
 - 9: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1988.DAT.*
 - 10: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1989.DAT.*
 - 11: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1990.DAT.*
 - 12: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1991.DAT.*
 - 13: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1992.DAT.*
 - 14: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1993.DAT.*
 - 15: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1994.DAT.*
 - 16: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1995.DAT.*
 - 17: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1996.DAT.*
 - 18: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1997.DAT.*
 - 19: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1998.DAT.*
 - 20: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1999.DAT.*
 - 21: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2000.DAT.*
 - 22: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT.*
 - 23: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT.*
 - 24: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	12	100.0	12	21	AAA12002
2	12	100.0	24	21	AAA12003
3	12	100.0	35	15	AAQ78668
4	12	100.0	36	15	AAQ78675
5	12	100.0	159	20	AAQ89029
6	12	100.0	170	10	AAQ90231
7	12	100.0	174	15	AAQ78658
8	12	100.0	177	9	AAQ80185
9	12	100.0	185	9	AAQ81966

C 10	12	100.0	185	10	AAQ90230	Epidermal growth f
C 11	12	100.0	191	10	AAQ91199	Chemically synthe
C 12	12	100.0	237	21	AAA12001	Murine PGK HRE der
C 13	12	100.0	330	13	AAQ20579	[Phe23]EGF in pTA
C 14	12	100.0	346	21	AAA44708	Human secreted exp
C 15	12	100.0	515	10	AAQ92328	Synthetic human ep
C 16	12	100.0	543	20	AAQ89038	EGF-angiogenin fus
C 17	12	100.0	554	22	AAQ88525	Human digestive sy
C 18	12	100.0	558	20	AAQ89039	Human angiogenin fus
C 19	12	100.0	575	24	AAQ63085	EGF-angiogenin fus
C 20	12	100.0	617	21	AAQ44867	Rat sequence diffe
C 21	12	100.0	717	20	AAQ89045	Arabidopsis thalia
C 22	12	100.0	762	22	AAQ85250	EGF-angiogenin fus
C 23	12	100.0	771	21	AAQ15094	Human cDNA clone (
C 24	12	100.0	825	21	AAQ41178	Trichoderma reesei
C 25	12	100.0	927	20	AAQ89044	Arabidopsis thalia
C 26	12	100.0	953	16	AAQ80299	EGF-angiogenin fus
C 27	12	100.0	953	22	AAQ91533	Plasmid insert com
C 28	12	100.0	953	22	AAQ91541	Miflag-EGF-pIII fu
C 29	12	100.0	975	23	AAQ94402	DNA encoding novel
C 30	12	100.0	984	21	AAQ44788	Zea mays DNA fragm
C 31	12	100.0	1014	22	AAQ52381	Human polynucleoti
C 32	12	100.0	1014	22	AAQ53365	Human polynucleoti
C 33	12	100.0	1134	23	AAQ94405	Human polynucleoti
C 34	12	100.0	1275	23	AAQ90258	DNA encoding novel
C 35	12	100.0	1351	22	AAQ64804	DNA encoding novel
C 36	12	100.0	1380	21	AAQ54172	Human secreted pro
C 37	12	100.0	1380	21	AAQ54173	Neisseria meningit
C 38	12	100.0	1399	23	ABQ09973	Drosophila melanog
C 39	12	100.0	1474	22	AAQ64863	Human secreted pro
C 40	12	100.0	1572	24	AAQ68446	Streptococcus poly
C 41	12	100.0	1677	22	AAQ46938	DNA encoding novel
C 42	12	100.0	1725	23	AAQ90262	DNA encoding novel
C 43	12	100.0	1725	23	AAQ94413	Secreted protein g
C 44	12	100.0	1781	22	AAQ72817	C glutamic codin
C 45	12	100.0	1845	22	AAQ68452	

ALIGNMENTS

RESULT 1
AAA12002
ID AAA12002 standard; DNA; 12 BP.
XX
AC AAA12002;
XX
DT 14-AUG-2000 (first entry)
XX
DE Murine HIF-1 spacer DNA #1.
XX

XX HRE; hypoxia response element; hypoxia-inducible factor; HIF; vasotropic;
XX cardiant; cytostatic; antithratic; gene therapy; ischaemia; arthritis;
XX cardiovascular disease; peripheral arterial disease; cancer; murine; ds.
XX Mus sp.
XX WO200017371-A1.
XX
XX 30-MAR-2000.
XX
XX 22-SEP-1999; 99WO-GB03181.
XX
XX 23-SEP-1998; 98WO-GB02885.
XX
XX 28-JAN-1999; 99GB-0001906.
XX
XX 16-FEB-1999; 99GB-0003538.
XX
XX (OXFO-) OXFORD BIOMEDICA UK LTD.
XX
XX Binley KM, Naylor S;
XX
XX WPI; 2000-283595/24.
XX

Novel polynucleotide constructs comprising at least two repeats of a hypoxia response element useful for driving expression of nucleic acids of interest in a cell under hypoxic conditions

Disclosure; Page 10; 155pp; English.

This invention describes novel polynucleotide comprising at least 2 repeats of a hypoxia response element (HRE), where the hypoxia-inducible factor (HIF) consensus binding sites within each of the 2 repeats are separated by a spacer of at least 20 contiguous nucleotides. The products of the invention have vasotropic, cardiant, cytostatic and antiarthritic activity and can be used for gene therapy. The polynucleotides are useful for delivering nucleic acids of interest to mammalian cells. Lentiviral vectors are responsive to hypoxic agents and to agents that mimic hypoxia. This regulation can be harnessed in vivo to enhance the production of the vector and can be used in vivo to regulate gene expression in response to a physiological signal. The vectors have utility in disease, where ischaemia, including hypoxia, is a feature, e.g. cardiovascular disease, peripheral arterial disease, cancer and arthritis. The novel regulatory construct is capable of driving very high levels of transcription under conditions of hypoxia whilst providing only low basal levels of transcription under normal oxygen conditions. The polynucleotide construct targets cells within a tumor mass that are under conditions of hypoxia without affecting normal surrounding tissue. This sequence represents a murine HIF-1 DNA spacer which is used in the method of the invention.

Sequence 12 BP; 2 A; 3 C; 5 G; 2 T; 0 other;

Query Match 100.0%; Score 12; DB 21; Length 12;
Best Local Similarity 100.0%; Pred. No. 1.9e+03;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GTCGTGCAGGCA 12
| | | | | | | | | | | | | |
Db 1 GTCGTGCAGGCA 12

RESULT 2
AAAL2003
ID AAA12003 standard; DNA; 24 BP.
XX
AC AAAL2003;
XX
XX 14-AUG-2000 (first entry)
XX
DE Murine HIF-1 spacer DNA #2.
XX
KW HRE; hypoxia response element; hypoxia-inducible factor; HIF; vasotropic;
KW cardiant; cytostatic; antiarthritic; gene therapy; ischaemia; arthritis;
KW cardiovascular disease; peripheral arterial disease; cancer; murine; ds.
XX
OS Mus sp.
XX
PN WO200017371-A1.
XX
XX 30-MAR-2000.
XX
PF 22-SEP-1999; 99WO-GB03181.
XX
PR 23-SEP-1998; 98WO-GB02885.
PR 28-JAN-1999; 99GB-0001906.
PR 16-FEB-1999; 99GB-0003538.
XX
XX (OXFO-) OXFORD BIOMEDICA UK LTD.
XX
XX Binley KM, Naylor S;
PI WPI; 2000-283595/24.
DR
XX Novel polynucleotide constructs comprising at least two repeats of a
PT hypoxia response element useful for driving expression of nucleic acids
PT of interest in a cell under hypoxic conditions

Novel polynucleotide comprising at least 2 repeats of a hypoxia response element (HRE), where the hypoxia-inducible factor (HIF) consensus binding sites within each of the 2 repeats are separated by a spacer of at least 20 contiguous nucleotides. The products of the invention have vasotropic, cardiant, cytostatic and antiarthritic activity and can be used for gene therapy. The polynucleotides are useful for delivering nucleic acids of interest to mammalian cells. Lentiviral vectors are responsive to hypoxic agents and to agents that mimic hypoxia. This regulation can be harnessed in vivo to enhance the production of the vector and can be used in vivo to regulate gene expression in response to a physiological signal. The vectors have utility in disease, where ischaemia, including hypoxia, is a feature, e.g. cardiovascular disease, peripheral arterial disease, cancer and arthritis. The novel regulatory construct is capable of driving very high levels of transcription under conditions of hypoxia whilst providing only low basal levels of transcription under normal oxygen conditions. The polynucleotide construct targets cells within a tumor mass that are under conditions of hypoxia without affecting normal surrounding tissue. This sequence represents a murine HIF-1 DNA spacer which is used in the method of the invention.

Sequence 24 BP; 4 A; 5 C; 7 G; 8 T; 0 other;

Query Match 100.0%; Score 12; DB 21; Length 24;
Best Local Similarity 100.0%; Pred. No. 1.9e+03;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GTCGTGCAGGCA 12
| | | | | | | | | | | | | |
Db 7 GTCGTGCAGGCA 18

RESULT 3
AAQ78668/c
ID AAQ78668 standard; DNA; 35 BP.
XX
AC AAQ78668;
XX
XX 20-JUL-1995 (first entry)
XX
DE Human epidermal growth factor DNA construction oligo, C2.
XX
KW Human; epidermal growth factor; hEGF; expression; E. coli;
KW OmpA leader sequence; universal translation termination;
KW trpA transcription termination; secretion; ss.
XX
OS Synthetic.
XX
PN WO9425592-A.
XX
XX 10-NOV-1994.
XX
XX 25-APR-1994; 94WO-KR00036.
XX
PR 26-APR-1993; 93KR-0006978.
PR 26-APR-1993; 93KR-0006979.
PR 26-APR-1993; 93KR-0006980.
XX
XX (DAEW-) DAEWOONG PHARM CO LTD.
XX
XX Chung JY, Jee YS, Koh YW, Kwon CH, Lee KM, Nho KS;
PI Park SK, Yu YH;
XX
XX WPI; 1994-358269/44.
XX
XX Novel gene encoding human epidermal growth factor - useful for
PT transformation of E.coli to give high yield expression.
XX
XX Example 2; Fig 2a; 40pp; English.

CC The sequences given in AAQ78667-76 are oligonucleotides which, when
 CC ligated together form a human epidermal growth factor (hEGF) coding
 CC sequence which comprises a HpaI restriction site at its 5' terminal, a
 CC PstI site at the 3' terminal and Bpu102I, NsiI, MluI, Eco47III and
 CC AflII restriction sites within the coding sequence. The hEGF contains
 CC codon usage which is biased for expression in E. coli. This sequence
 CC may be inserted into a OmpA leader-universal translation termination-
 CC trpA transcription termination sequence between the OmpA leader and
 CC universal translation termination sequence. This full length sequence
 CC may then be used for the expression and secretion of hEGF in E. coli.
 XX
 SQ Sequence 35 BP; 7 A; 11 C; 10 G; 7 T; 0 other;

Query Match 100.0%; Score 12; DB 15; Length 35;
 Best Local Similarity 100.0%; Pred. No. 1.9e+03;
 Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GTCGTGCAGGCA 12
 |||||
 Db 24 GTCGTGCAGGCA 13

RESULT 4
 AAQ78675
 ID AAQ78675 standard; DNA; 36 BP.
 AC AAQ78675;

DT 20-JUL-1995 (first entry)

DE Human epidermal growth factor DNA construction oligo, M4.

KW Human; epidermal growth factor; hEGF; expression; E. coli;
 KW OmpA leader sequence; universal translation termination;
 KW trpA transcription termination; secretion; ss.

OS Synthetic.

PN WO9425592-A.

PD 10-NOV-1994.

PF 25-APR-1994; 94WO-KR00036.

PR 26-APR-1993; 93KR-0006978.

PR 26-APR-1993; 93KR-0006979.

PR 26-APR-1993; 93KR-0006980.

PA (DAEW-) DAEWOONG PHARM CO LTD.

PI Chung JY, Jee YS, Koh YW, Kwon CH, Lee KM, Nho KS;
 PI Park SK, Yu YH;

DR WPI; 1994-358269/44.

PT Novel gene encoding human epidermal growth factor - useful for
 PT transformation of E.coli to give high yield expression.

PS Example 2; Fig 2a; 40pp; English.

CC The sequences given in AAQ78667-76 are oligonucleotides which, when
 CC ligated together form a human epidermal growth factor (hEGF) coding
 CC sequence which comprises a HpaI restriction site at its 5' terminal, a
 CC PstI site at the 3' terminal and Bpu102I, NsiI, MluI, Eco47III and
 CC AflII restriction sites within the coding sequence. The hEGF contains
 CC codon usage which is biased for expression in E. coli. This sequence
 CC may be inserted into a OmpA leader-universal translation termination-
 CC trpA transcription termination sequence between the OmpA leader and
 CC universal translation termination sequence. This full length sequence
 CC may then be used for the expression and secretion of hEGF in E. coli.

XX Sequence 36 BP; 8 A; 9 C; 11 G; 8 T; 0 other;

Query Match 100.0%; Score 12; DB 15; Length 36;
 Best Local Similarity 100.0%; Pred. No. 1.9e+03;
 Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GTCGTGCAGGCA 12
 |||||
 Db 21 GTCGTGCAGGCA 32

RESULT 5

ID AAX89029/C
 XX AAX89029 standard; DNA; 159 BP.

AC AAX89029;

DT 15-SEP-1999 (first entry)

DE Human epidermal growth factor (hEGF) encoding DNA.

KW Fusion protein; epidermal growth factor; hEGF; human; angiogenin;
 KW anticancer agent; cancer; cytotoxicity; ss.

OS Homo sapiens.

PN WO9923112-A1.

PD 14-MAY-1999.

PF 30-OCT-1998; 98WO-KR00343.

PR 01-NOV-1997; 97KR-0057603.

PA (DAEW-) DAEWOONG PHARM CO LTD.

PI Han S, Kim S, Kim Y, Koo T, Kwon O, Lee B, Park M;
 PI Park S, Yoon J;

DR WPI; 1999-418417/35.

PT P-PSDB; AAY27102.

PT Human epidermal growth factor and human angiogenin fusion proteins
 Example 1; Page 36; 73pp; English.

CC The invention provides a fusion protein consisting of human epidermal
 CC growth factor (hEGF) and human angiogenin joined by a linker. The hEGF-
 CC angiogenin fusion protein is useful as an anticancer agent. The hEGF-
 CC tracks down cancer cells expressing hEGF receptors at high level
 CC following internalisation, and the angiogenin exhibits cytotoxicity by
 CC degrading ribonucleic acids upon internalisation. The hEGF-angiogenin
 CC fusion consisting of hEGF and angiogenin, both of which normally exist in
 CC the human body, exhibit no toxicity following overdose administration in
 CC bacteria. The fusion protein selectively inhibits the growth of cancer
 CC cells expressing the hEGF receptor. It does not have a detrimental effect
 CC on the growth of the normal cells. Further it does not exhibit toxicity
 CC of conventional chemical anti-cancer agents and it does not cause any
 CC serious problem by forming antibody against the fusion protein. Targeting
 CC efficiency is improved using small molecular weight proteins, which are
 CC 6 kD hEGF and 14.4 kD angiogenin. The present sequence represents a DNA
 CC encoding a hEGF.

SQ Sequence 159 BP; 36 A; 42 C; 47 G; 34 T; 0 other;

Query Match 100.0%; Score 12; DB 20; Length 159;
 Best Local Similarity 100.0%; Pred. No. 1.8e+03;
 Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GTCGTGCAGGCA 12
 |||||
 Db 51 GTCGTGCAGGCA 40

Mon Jan 6 15:20:18 2003

us-09-787-562-10.rng

```

RESULT 6
AA090231/c
ID AA090231 standard; DNA; 170 BP.
XX
XX
XX AC
XX 01-NOV-1989 (first entry)
XX
XX Alternative sequence of epidermal growth factor structural gene.
DE
XX
XX Recombinant plasmid pAMPUG; epidermal growth factor; ampicillin
KW resistance; Escherichia coli.
XX
XX Key Location/Qualifiers
FH misc_feature 11..17
FT /*tag= a
FT CDS 18..164
FT /*tag= b
FT misc_feature 88..91
FT /*tag= c
XX
XX JP01132383-A.
XX
XX 24-MAY-1989.
XX
XX 18-NOV-1987; 87JP-0289332.
XX
XX 18-NOV-1989; 89JP-0289332.
XX
XX (HITA ) HITACHI KK.
PA (HITB ) HITACHI CHEMICAL KK.
XX
XX WPI; 1989-195592/27.
DR P-PSDB; AAP90465.
XX
XX Recombinant plasmid pAMPUG with epidermal growth activity - comprises the
PT epidermal growth factor structural gene joined to HincII site of
PT ampicillin resistance gene.
XX
XX PS Disclosure; fig 2; 4pp; Japanese.
XX
XX The sequence is an alternative sequence of epidermal growth factor
CC structural gene (see AA090230), which is ligated to the HincII site of
CC ampicillin resistance gene, and inserted into recombinant plasmid
CC pAMPUG. One or more strong promoter is used, eg tac, lac or trp promoter.
CC esp. the amp resistance gene promoter and Shine Dalgarno sequence. The
CC EGF is expressed in E.coli. Misc. feature a is an XbaI site, and c is a
CC TaqI site. See also AAP90465.
XX
XX SQ Sequence 170 BP; 44 A; 41 C; 45 G; 40 T; 0 other;

Query Match 100.0%; Score 12; DB 10; Length 170;
Best Local Similarity 100.0%; Pred. No. 1.8e+03;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GTCGTCGAGGCA 12
|||||
DB 71 GTCGTCGAGGCA 60

RESULT 7
AAQ78658/c
ID AAQ78658 standard; DNA; 174 BP.
XX
XX
XX AC
XX
XX 20-JUL-1995 (first entry)
XX
XX Epidermal growth factor DNA sequence.
DE
XX
XX Human; epidermal growth factor; hEGF; expression; E. coli;
KW OmpA leader sequence; universal translation termination;
KW trpA transcription termination; secretion; ss.

```

```

XX OS Homo sapiens.
XX
XX PN W09425592-A.
XX
XX PD 10-NOV-1994.
XX
XX PF 25-APR-1994; 94WO-KR00036.
XX
XX PR 26-APR-1993; 93KR-0006978.
XX
XX PR 26-APR-1993; 93KR-0006979.
XX
XX PR 26-APR-1993; 93KR-0006980.
XX
XX (DAEW-) DAEWOONG PHARM CO LTD.
XX
XX PI Chung JY, Jee YS, Koh YW, Kwon CH, Lee KM, Nho KS;
PI Park SK, Yu YH;
XX
XX WPI; 1994-358269/44.
XX
XX Novel gene encoding human epidermal growth factor - useful for
PT transformation of E.coli to give high yield expression.
XX
XX Claim 1; Page 21; 40pp; English.
XX
XX This sequence represents the coding sequence of a human epidermal
CC growth factor which comprises a HpaI restriction site at the 5'
CC terminal, a PstI site at the 3' terminal and Bpu102I, NsiI, MluI,
CC Eco47III and AflII restriction sites within the coding sequence.
CC This sequence encodes a naturally occurring epidermal growth factor,
CC but the codon usage is biased for expression in E. coli. This
CC sequence may be inserted into a OmpA leader-universal translation
CC termination-trpA transcription termination sequence between the OmpA
CC leader and universal translation termination sequence. This
CC sequence may then be used for the expression and secretion of
CC epidermal growth factor in E. coli.
XX
XX SQ Sequence 174 BP; 40 A; 46 C; 51 G; 37 T; 0 other;

Query Match 100.0%; Score 12; DB 15; Length 174;
Best Local Similarity 100.0%; Pred. No. 1.8e+03;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GTCGTCGAGGCA 12
|||||
DB 54 GTCGTCGAGGCA 43

RESULT 8
AA080185/c
ID AA080185 standard; DNA; 177 BP.
XX
XX AC
XX
XX DT 14-NOV-1990 (first entry)
XX
XX DNA encoding biosynthetic multifunctional protein.
DE
XX
XX KW Biosynthetic multifunctional protein; biosynthetic antibody binding site;
CC protein trailer; epidermal growth factor; ss DNA.
XX
XX FH Key Location/Qualifiers
FT CDS 1..171
FT /*tag= a
XX
XX PN W08809344-A.
XX
XX PD 01-DEC-1988.
XX
XX PF 19-MAY-1988; 88WO-US01737.
XX
XX PR 21-MAY-1987; 87US-0052800.
XX
XX

```

```
PA (CREA-) CREATIVE BIOMOLECULES INC.
XX
XX Huston JS, Oppermann H;
XX
XX WPI: 1988-353928/49.
XX P-PSDB; AAP80159.
XX
XX Recombinant multifunctional protein - having antibody binding site and a
XX sequence for biological activity, ion sequestering or binding to a
XX solid support.
XX
XX Disclosure; : 115pp; English.
XX
XX The DNA encodes a biosynthetic multifunctional protein including a single
XX chain biosynthetic antibody binding site and an epidermal growth factor
XX protein trailer.
XX
XX Sequence 177 BP; 39 A; 45 C; 53 G; 40 T; 0 other;
XX
XX Query Match 100.0%; Score 12; DB 9; Length 177;
XX Best Local Similarity 100.0%; Pred. No. 1.8e+03;
XX Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX QY 1 GTCGTGCAGGCA 12
XX |||||
XX Db 60 GTCGTGCAGGCA 49
XX
XX RESULT 9
XX AAN81966/c
XX ID AAN81966 standard; DNA; 185 BP.
XX
XX AC AAN81966;
XX
XX DT 09-OCT-1990 (first entry)
XX
XX DE Synthetic epidermal growth factor gene.
XX
XX KW Epidermal growth factor; ds.
XX
XX OS Synthetic.
XX
XX FH Key Location/Qualifiers
XX CDS 18..185
XX FT /*tag= a
XX FT /product=EGF
XX FT 6..9
XX FT /*tag= b
XX
XX PN JP63003790-A.
XX
XX PD 08-JAN-1988.
XX
XX PF 25-JUN-1986; 86JP-0146963.
XX
XX PR 25-JUN-1986; 86JP-0146963.
XX
XX PA (HITA ) HITACHI KK.
XX (HITB ) HITACHI CHEMICAL KK.
XX
XX WPI: 1988-046256/07.
XX P-PSDB; AAP81522.
XX
XX Recombinant plasmid with higher prodn. efficiency - is prepd. by
XX inserting a structural gene, encoding a 53-residue polypeptide,
XX downstream of tac promoter.
XX
XX Disclosure; : ; Japanese.
XX
XX The sequence was synthesized from 40 oligonucleotide fragments. The
XX 5' end of the sense strand overhangs the complementary strand by 4
XX bases; the 5' end of the complementary strand overhangs the 3' end of
XX the sense strand by -CTAG. The sequence is also given in J633791-A.
XX
XX Sequence 185 BP; 46 A; 44 C; 52 G; 43 T; 0 other;
XX
XX Query Match 100.0%; Score 12; DB 10; Length 185;
XX Best Local Similarity 100.0%; Pred. No. 1.8e+03;
XX Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX QY 1 GTCGTGCAGGCA 12
XX |||||
XX Db 71 GTCGTGCAGGCA 60
XX
XX RESULT 10
XX AAN90230/c
XX ID AAN90230 standard; DNA; 185 BP.
XX
XX AC AAN90230;
XX
XX DT 01-NOV-1989 (first entry)
XX
XX DE Epidermal growth factor structural gene.
XX
XX KW Recombinant plasmid pAMPUG; ampicillin resistance;
XX Escherichia coli.
XX
XX FH Key Location/Qualifiers
XX CDS 11..16
XX FT /*tag= a
XX FT 18..179
XX FT /*tag= b
XX FT 88..91
XX FT /*tag= c
XX
XX PN JF01132383-A.
XX
XX PD 24-MAY-1989.
XX
XX PF 18-NOV-1987; 87JP-0289332.
XX
XX PR 18-NOV-1989; 89JP-0289332.
XX
XX (HITA ) HITACHI KK.
XX (HITB ) HITACHI CHEMICAL KK.
XX
XX WPI: 1989-195592/27.
XX P-PSDB; AAP90465.
XX
XX Recombinant plasmid pAMPUG with epidermal growth activity - comprises the
XX epidermal growth factor structural gene joined to HincII site of
XX ampicillin resistance gene.
XX
XX Disclosure; fig 2; 4pp; Japanese.
XX
XX The sequence is of epidermal growth factor structural gene, which is
XX ligated to the HincII site of ampicillin resistance gene, and inserted
XX into recombinant plasmid pAMPUG. One or more strong promoter is used,
XX eg tac, lac or trp promoter, esp. the amp resistance gene promoter and
XX Shine Dalgarno sequence. The EGF is expressed in E.coli (see AAP90464
XX for encoded peptide). Misc. feature a is an XbaI site, and c is a TaqI
XX site. See also AAN90231 and AAP90465.
XX
XX Sequence 185 BP; 46 A; 44 C; 52 G; 43 T; 0 other;
XX
XX Query Match 100.0%; Score 12; DB 10; Length 185;
XX Best Local Similarity 100.0%; Pred. No. 1.8e+03;
XX Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX QY 1 GTCGTGCAGGCA 12
XX |||||
XX Db 71 GTCGTGCAGGCA 60
XX
XX RESULT 11
```

```

AAN91199/c
ID AAN91199 standard; DNA; 191 BP.
XX AC AAN91199;
XX DT 05-FEB-1990 (first entry)
XX DE Chemically synthesised human epidermal growth factor (hEGF) gene.
XX KW Human epidermal growth factor (hEGF); pTREBT; pTRLBT
XX KW PBR322; pTREBT; pTRLBT
XX FH Key Location/Qualifiers
XX FT CDS 24..191
XX FT /*tag= a
XX PN EP335400-A.
XX PD 04-OCT-1989.
XX PF 30-MAR-1989; 89EP-0105639.
XX PR 30-MAR-1988; 88JP-0074383.
XX PA (HITA ) HITACHI KK.
XX PI Shimizu N, Harada Y, Fukuzono S, Fujimori K;
XX DR WPI; 1989-286891/40.
XX DR P-PSDB;P9187.
XX PT DNA contg. human epidermal growth factor gene
XX PT - associated with E. coli cryptophan operon components
XX PS Claim 1; figure 1; 26pp; English.
XX CC Entire sequence of organically synthesised hEGF gene. Has 5' AATT single-
XX CC stranded sticky end on coding strand and 5' CTAG single-stranded sticky
XX CC end on non-coding strand.
XX SQ Sequence 191 BP; 48 A; 45 C; 53 G; 45 T; 0 other;

Query Match 100.0%; Score 12; DB 10; Length 191;
Best Local Similarity 100.0%; Pred. No. 1.8e+03;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GTCGTCGAGGCA 12
DB 77 GTCGTCGAGGCA 66

RESULT 12
AAL12001
ID AAL12001 standard; DNA; 237 BP.
XX AC AAL12001;
XX DT 14-AUG-2000 (first entry)
XX DE Murine PGK HRE derived promoter OBHrell DNA.
XX KW HRE; hypoxia response element; hypoxia-inducible factor; HIF; vasotrophic;
XX KW cardiant; cytosstatic; antiarthritic; gene therapy; ischaemia; arthritis;
XX KW cardiovascular disease; peripheral arterial disease; cancer;
XX KW phosphoglycerate kinase; PGK; murine; promoter; OBHrell; ds.
XX OS Mus sp.
XX PN WO200017371-A1.
XX PD 30-MAR-2000.
XX PF 22-SEP-1999; 99WO-GB03181.

```

```

XX 23-SEP-1998; 98WO-GB02885.
XX 28-JAN-1999; 99GB-0001906.
XX 16-FEB-1999; 99GB-0003538.
XX (OXFO-) OXFORD BIOMEDICA UK LTD.
XX PI Binley KM, Naylor S;
XX DR WPI; 2000-283595/24.
XX PT Novel polynucleotide constructs comprising at least two repeats of a
XX FT hypoxia response element useful for driving expression of nucleic acids
XX FT of interest in a cell under hypoxic conditions
XX PS Example 1; Page 68; 155pp; English.
XX CC This invention describes novel polynucleotide comprising at least 2
XX CC repeats of a hypoxia response element (HRE), where the hypoxia-inducible
XX CC factor (HIF) consensus binding sites within each of the 2 repeats are
XX CC separated by a spacer of at least 20 contiguous nucleotides. The products
XX CC of the invention have vasotropic, cardiant, cytosstatic and antiarthritic
XX CC activity and can be used for gene therapy. The polynucleotides are useful
XX CC for delivering nucleic acids of interest to mammalian cells. Lentiviral
XX CC vectors are responsive to hypoxic agents and to agents that mimic
XX CC hypoxia. This regulation can be harnessed in vitro to enhance the
XX CC production of the vector and can be used in vivo to regulate gene
XX CC expression in response to a physiological signal. The vectors have
XX CC utility in disease, where ischaemia, including hypoxia, is a feature,
XX CC e.g. cardiovascular disease, peripheral arterial disease, cancer and
XX CC arthritis. The novel regulatory construct is capable of driving very high
XX CC levels of transcription under conditions of hypoxia whilst providing only
XX CC low basal levels of transcription under normal oxygen conditions. The
XX CC polynucleotide construct targets cells within a tumor mass that are under
XX CC conditions of hypoxia without affecting normal surrounding tissue. This
XX CC sequence represents a murine phosphoglycerate kinase (PGK) HRE derived
XX CC promoter OBHrell which is described in the method of the invention.
XX SQ Sequence 237 BP; 43 A; 82 C; 56 G; 56 T; 0 other;

Query Match 100.0%; Score 12; DB 21; Length 237;
Best Local Similarity 100.0%; Pred. No. 1.8e+03;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GTCGTCGAGGCA 12
DB 31 GTCGTCGAGGCA 42

RESULT 13
AAQ20579/c
ID AAQ20579 standard; DNA; 330 BP.
XX AC AAQ20579;
XX DT 09-JUN-1992 (first entry)
XX DE [Phe23]EGF in pTA 1522-Eco.
XX KW Epidermal growth factor; variant; mutant; receptor; anticancer; ds.
XX OS Homo sapiens.
XX FH Key Location/Qualifiers
XX FT CDS 95..319
XX FT /*tag= a
XX FT /label= EGF
XX FT repeat_unit 4..25
XX FT /*tag= b
XX FT /rpt_type= inverted
XX FT repeat_unit 28..41
XX FT /*tag= c
XX FT /rpt_type= inverted

```

FT repeat_unit 44..111
 FT /*tag= d
 FT /rpt_type= inverted
 FT misc_signal 43..49
 FT /*tag= e
 FT /label= PB
 FT RBS 83..86
 FT /*tag= f
 FT sig_peptide 95..157
 FT /*tag= g
 FT /label= phoA_signal
 FT /note= "21 amino acids"
 FT mat_peptide 158..319
 FT /*tag= h
 FT /label= EGF
 FT misc_difference 224..226
 FT /*tag= i
 FT /note= "mutated to encode Phe instead of Ile"
 FT JP03294293-A.
 PN 25-DEC-1991.
 PD 11-APR-1990; 90JP-0093921.
 PF 11-APR-1990; 90JP-0093921.
 PR (YUEI-) YUEI SEIYAKU KK.
 XX WPI: 1992-052025/07.
 DR P-PSDB; AAR20623.
 XX

Human epithelial cell growth factor variant - with higher
 receptor activity than natural EGF, useful in development of
 anticancer drugs
 PS Disclosure; Fig 1; 6pp; Japanese.

CC pTA1522-Eco comprises a phoA promoter-, EGF- and a phoA- signal
 CC sequence. The codon for Ile at position 23 is changed to TTT
 CC (encoding Phe) by oligonucleotide site-directed mutagenesis to
 CC produce the new variant. The variant has a higher affinity for
 CC receptors than natural hEGF and may be used in the development of
 CC anticancer drugs.

XX SQ Sequence 330 BP; 83 A; 72 C; 78 G; 97 T; 0 other;

Query Match 100.0%; Score 12; DB 13; Length 330;
 Best Local Similarity 100.0%; Pred. No. 1.8e+03;
 Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GTCGTGCAGGCA 12
 |||||
 Db 208 GTCGTGCAGGCA 197

RESULT 14

AAA44708
 ID AAA44708 standard; cDNA; 346 BP.

XX
 AC AAA44708;

XX 21-AUG-2000 (first entry)

XX Human secreted expressed sequence tag SEQ ID NO:1283.

XX Human; mouse; chicken; rat; secreted expressed sequence tag; SEST;
 XX expressed sequence tag; EST; probe; chemotactic; proliferative;
 KW immunomodulatory; haematopoietic; chemokinetic; analgesic; haemostatic;
 KW thrombolytic; antiinflammatory; cytostatic; antibacterial; antifungal;
 KW antiviral; antidiabetic; antiasthmatic; vulnery; antiparkinsonian;
 KW antiulcer; osteopathic; neuroprotective; nootropic; antipsoriatic;
 KW cerebroprotective; anticonvulsant; antidepressant; gene therapy;

KW vaccine; autoimmune disorder; multiple sclerosis; allergic condition;
 KW insulin dependent diabetes; asthma; myeloid cell deficiency; ulcer;
 KW lymphoid cell deficiency; burn; osteoporosis; osteoarthritis;
 KW central nervous system disorder; Alzheimer's disease; stroke;
 KW Parkinson's disease; Huntington's disease; coagulation disorder;
 KW haemophilia; thrombosis; inflammatory disorder; Crohn's disease;
 KW tumour; infection; depression; psoriasis; ss.

OS Homo sapiens.

XX WO200021991-A1.

XX 20-APR-2000.

PF 15-OCT-1999; 99WO-US24206.

PR 15-OCT-1998; 98US-0104436.

XX (GEMY) GENETICS INST INC.

XX Jacobs K, McCoy JM, LaVallie ER, Collins-Racie LA, Evans C;
 PI Merberg D, Treacy M, Bowman MR;
 XX WPI: 2000-317938/27.

XX Isolated polynucleotides, and encoded proteins, comprising secreted
 PT expressed sequence tags (SESTs), useful for treating various disorders
 PT such as autoimmune, infectious, and central nervous system disorders -
 XX Claim 1; Page 535; 803pp; English.

XX AAA43426 to AAA45925 represent specifically claimed secreted expressed
 CC sequence tags (SESTs), isolated from human, mouse, chicken and rat
 CC tissue sources. The SESTs can have a range of activities depending on
 CC the tissues they were isolated from. The activities include:
 CC chemokinetic; proliferative; immunomodulatory; haematopoietic;
 CC cytostatic; antibacterial; antifungal; antiviral; antiinflammatory;
 CC antiasthmatic; vulnery; antiulcer; osteopathic; neuroprotective;
 CC nootropic; antiparkinsonian; antipsoriatic; cerebroprotective;
 CC anticonvulsant; and antidepressant. The SESTs can be used for gene
 CC therapy and in vaccines. The SESTs are useful as probes for the
 CC identification and isolation of full-length cDNAs and genomic DNA
 CC molecules which correspond to the SESTs. Proteins encoded by the SESTs
 CC are useful in assays for determining biological activity and raising
 CC antibodies. They may be useful for treatment of autoimmune disorders
 CC (multiple sclerosis, insulin dependent diabetes), allergic conditions
 CC (asthma), myeloid or lymphoid cell deficiencies, wounds, ulcers,
 CC osteoporosis, osteoarthritis, central nervous system disorders
 CC (Alzheimer's, Parkinson's, Huntington's disease, stroke), coagulation
 CC disorders (haemophilia, thrombosis), inflammatory disorders (Crohn's
 CC disease), tumours, bacterial, fungal or viral infections, depression and
 CC psoriasis. AAA45926 to AAA45931 represent linker variants which are given
 CC in the exemplification of the present invention.

SQ Sequence 346 BP; 79 A; 61 C; 131 G; 75 T; 0 other;

Query Match 100.0%; Score 12; DB 21; Length 346;
 Best Local Similarity 100.0%; Pred. No. 1.8e+03;

Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GTCGTGCAGGCA 12

|||||
 Db 87 GTCGTGCAGGCA 98

RESULT 15

AAN92328/C

ID AAN92328 standard; DNA; 515 BP.

XX AC AAN92328;

XX DT 25-APR-1990 (first entry)

```

XX Synthetic human epidermal growth factor.
DE
XX Human epidermal growth factor; EGF;
XX
XX Key Location/Qualifiers
XX CDS 18..570
XX /*tag= a
XX repeat_unit 171..350
XX /*tag= b
XX
XX JP01257482-A.
XX
XX 13-OCT-1989.
XX
XX 08-APR-1988; 88JP-0085073.
XX PF
XX 08-APR-1988; 88JP-0085073.
XX PR
XX (HITA ) HITACHI KK.
XX (HITB ) HITACHI CHEMICAL KK.
XX
XX WPI; 1989-344714/47.
XX P-PSDB; AAP93397.
XX
XX DNA contg. synthetic gene - used for human epidermal growth factor prodn.
XX
XX Disclosure; fig. 2; 7pp; Japanese.
XX
XX The gene has the following structure: tag b:(AAPi,m)n:
XX Tag b is a repeating unit encoding hEGF. The synthetic gene
XX can be inserted into a plasmid vector for efficient prodn. of EGF.
XX
XX Sequence 515 BP; 125 A; 125 C; 148 G; 117 T; 0 other;
SQ
Query Match 100.0%; Score 12; DB 10; Length 515;
Best Local Similarity 100.0%; Pred. No. 1.8e+03;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GTCGTGCAGGCA 12
Db 71 GTCGTGCAGGCA 60

```

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 3, 2003, 22:54:17 ; Search time 159.558 Seconds
(without alignments)
1218.024 Million cell updates/sec

Title: US-09-787-562-10
Perfect score: 12
Sequence: 1 gtcgtgcaggca 12

Scoring table: IDENTITY_NUC
Gapop 10.0, Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST.*

- 1: em_estba.*
- 2: em_esthum.*
- 3: em_estnu.*
- 4: em_estnu.*
- 5: em_estov.*
- 6: em_estpl.*
- 7: em_estro.*
- 8: em_hic.*
- 9: gb_est1.*
- 10: gb_est2.*
- 11: gb_hic.*
- 12: gb_est3.*
- 13: gb_est4.*
- 14: gb_est5.*
- 15: em_estfun.*
- 16: em_estom.*
- 17: gb_gss.*
- 18: em_gss_hum.*
- 19: em_gss_inv.*
- 20: em_gss_pln.*
- 21: em_gss_vrt.*
- 22: em_gss_fun.*
- 23: em_gss_mam.*
- 24: em_gss_mus.*
- 25: em_gss_other.*
- 26: em_gss_pro.*
- 27: em_gss_rod.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	12	100.0	137	12	BG050523 FM1_72_D0
2	12	100.0	174	12	BF837005 PM3-HT090
3	12	100.0	177	9	AI030525 UI-R-CO-1
4	12	100.0	180	10	BE150969 RC4-HT027
5	12	100.0	192	13	BI024277 CM3-MT029
6	12	100.0	212	13	BI034110 PM2-NN016

7	12	100.0	221	9	AA815339
8	12	100.0	239	12	BF760820
9	12	100.0	248	17	BH227331
10	12	100.0	253	13	BM145815
11	12	100.0	256	12	BF256224
12	12	100.0	261	12	BF146813
13	12	100.0	286	10	BB179675
14	12	100.0	292	14	BM820141
15	12	100.0	296	9	AT004345
16	12	100.0	297	12	BE766861
17	12	100.0	298	9	AA469407
18	12	100.0	300	9	AJ467211
19	12	100.0	300	9	AJ468211
20	12	100.0	305	10	BB491289
21	12	100.0	307	10	AA401951
22	12	100.0	316	10	BE604099
23	12	100.0	323	17	AZ577296
24	12	100.0	325	10	BE146189
25	12	100.0	325	12	BE863787
26	12	100.0	328	14	Z48448
27	12	100.0	330	9	AI867483
28	12	100.0	331	9	AV158985
29	12	100.0	332	9	AA741679
30	12	100.0	332	10	BE517981
31	12	100.0	336	14	BQ166106
32	12	100.0	340	9	AL816780
33	12	100.0	344	14	BQ164373
34	12	100.0	346	13	BI974238
35	12	100.0	357	9	AA666066
36	12	100.0	359	12	BF359469
37	12	100.0	361	10	AW878486
38	12	100.0	363	10	BE086558
39	12	100.0	365	9	AI007372
40	12	100.0	368	10	BE634826
41	12	100.0	369	14	BQ279572
42	12	100.0	373	17	AQ099324
43	12	100.0	375	12	BG013635
44	12	100.0	378	12	BG234973
45	12	100.0	384	9	AA146187

ALIGNMENTS

RESULT 1
BG050523
LOCUS
DEFINITION
FM1_72_D03_g1_A003 Floral-Induced Meristem 1 (FMI) Sorghum
PROPINQUUM CDNA, mRNA sequence.
ACCESSION
BG050523
VERSION
BG050523.1
KEYWORDS
EST.
SOURCE
Sorghum propinquum.
ORGANISM
Sorghum propinquum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
clade; Panicoideae; Andropogoneae; Sorghum.
REFERENCE
1 (bases 1 to 137)
Cordonnier-Pratt, M.-M., Gingle, A., Sudman, M., Marsala, C. and Pratt, L.H.
TITLE
An EST database from Sorghum: floral-induced meristems
JOURNAL
Unpublished (2000)
COMMENT
Contact: Cordonnier-Pratt MM
Laboratory for Genomics and Bioinformatics
The University of Georgia, Department of Plant Biology
Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA
Tel: 706 542 1860
Fax: 706 583 0210
Email: mmpratt@uga.edu
Sequences have been trimmed to exclude PolyA, vector and regions
below Phred quality 16. The threshold for highest quality sequence
is 20.
Seq primer: PolyTMix

High quality sequence start: 29
High quality sequence stop: 70
POLYA=No.

Location
1. .137

```

/db_xref="taxon:132711"

```

phlucscript II from Lambda zap II; Site_1: XhoI; Site_2: EcoRI; mature plants were placed in a growth chamber for 15 days with 16 hr darkness and 8 hr light (flowering is induced by short-day conditions); 16 days after being returned to the greenhouse under natural long days during late April/early May, meristems were harvested. The library was made from poly-A RNA in the cloning vector lambda zap II. Clones to be sequenced were prepared by mass excision."

returned to the greenhouse late April/early May, meristems were excised and a library was made from poly- γ -ZAP II. Clones to be excised were identified by mass excision."

Year	34 a	26 c	31 g	46 t
1990	100	100	100	100
1991	100	100	100	100
1992	100	100	100	100
1993	100	100	100	100
1994	100	100	100	100
1995	100	100	100	100
1996	100	100	100	100
1997	100	100	100	100
1998	100	100	100	100
1999	100	100	100	100
2000	100	100	100	100
2001	100	100	100	100
2002	100	100	100	100
2003	100	100	100	100
2004	100	100	100	100
2005	100	100	100	100
2006	100	100	100	100
2007	100	100	100	100
2008	100	100	100	100
2009	100	100	100	100
2010	100	100	100	100
2011	100	100	100	100
2012	100	100	100	100
2013	100	100	100	100
2014	100	100	100	100
2015	100	100	100	100
2016	100	100	100	100
2017	100	100	100	100
2018	100	100	100	100
2019	100	100	100	100
2020	100	100	100	100
2021	100	100	100	100
2022	100	100	100	100
2023	100	100	100	100
2024	100	100	100	100
2025	100	100	100	100
2026	100	100	100	100
2027	100	100	100	100
2028	100	100	100	100
2029	100	100	100	100
2030	100	100	100	100
2031	100	100	100	100
2032	100	100	100	100
2033	100	100	100	100
2034	100	100	100	100
2035	100	100	100	100
2036	100	100	100	100
2037	100	100	100	100
2038	100	100	100	100
2039	100	100	100	100
2040	100	100	100	100
2041	100	100	100	100
2042	100	100	100	100
2043	100	100	100	100
2044	100	100	100	100
2045	100	100	100	100
2046	100	100	100	100
2047	100	100	100	100
2048	100	100	100	100
2049	100	100	100	100
2050	100	100	100	100
2051	100	100	100	100
2052	100	100	100	100
2053	100	100	100	100
2054	100	100	100	100
2055	100	100	100	100
2056	100	100	100	100
2057	100	100	100	100
2058	100	100	100	100
2059	100	100	100	100
2060	100	100	100	100
2061	100	100	100	100
2062	100	100	100	100
2063	100	100	100	100
2064	100	100	100	100
2065	100	100	100	100

Clarity	Score	DB
100.0%;	Score 12;	DB
100.0%;	Pred. No.	3.5

AGGCA 12
|||||

AGGCA 17

005
T0909-191100-028-f09 HT0909 Hom
005
005 J GT:12189164

Phyla: Echinodermata; Chordata; Crani-
yota; Metazoa; Chordata; Crani-
lia; Eutheria; Primates; Catarr-

Neto, E., Garcia Correa, R., Verj
M A da Silva, W. Jr., Zago, M.

man, G.H.; Carvalho, A.F.; Matsukura, T.; de Oliveira, P.S.; Buche-
stein, A.; Prontani, P.P.

son, A.J.
run sequencing of the human tran

Natl. Acad. Sci. U.S.A. 97 (7)
2663

Laboratory of Cancer Genetics
 The Cancer Research

55-11-2704922

433-11-2707001
E-mail: asimpson@ludwig.org.br
sequence was derived from the

ect. This entry can be seen in
p://www.ludwig.org.br/scripts/g
0-028-f00et3=2000-11-19&t4=1)

primer: puc 18 forward
quality sequence start: 27
quality sequence stop: 173

Location/Qualifiers
1. .174

```
/db_xref="taxon:9606"  
/clone_lib="HT0909"
```


allows identification of the library of origin of a clone within the mixture. The subtracted library (UI-R-C0) was constructed as follows: PCR amplified cDNA inserts from a pool of UI-R-A1 and UI-R-E1 clones from which 3' ESTs had been derived was used as a driver in a hybridization with the pooled UI-R-A1 and UI-R-E1 library in the form of single-stranded circles. The remaining single-stranded circles (subtracted library) was purified by hydroxyapatite column chromatography, converted to double-stranded circles and electroporated into DH10B bacteria (Life Technologies) to generate the UI-R-C0 library. This procedure has been previously described (Ronald, Lennon and Soares, Genome Research 6: 791-806, 1996).

BASE COUNT 40 a 31 c 47 g 59 t
 ORIGIN

Query Match 100.0%; Score 12; DB 9; Length 177;
 Best Local Similarity 100.0%; Pred. No. 3.6e+04;
 Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GTCGTGCAGGCA 12
 |||||

Db 85 GTCGTGCAGGCA 96

RESULT 4
 BE150969

LOCUS BE150969 180 bp mRNA linear EST 21-JUN-2000
 DEFINITION RC4-HP0276-100300-015-f11 HT0276 Homo sapiens cDNA, mRNA sequence.
 ACCESSION BE150969
 VERSION BE150969.1 GI:8613690
 KEYWORDS EST.
 SOURCE human.

ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Euthera; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 180)
 Dias Neto, E., Garcia Correa, R., Verjovski-Almeida, S., Briones, M.R.,
 Nagai, M.A., da Silva, W. Jr., Zago, M.A., Bordin, S., Costa, F.F.,
 Goldman, G.H., Carvalho, A.F., Matsukuma, A., Baia, G.S., Simpson, D.H.,
 Brunstein, A., deOliveira, P.S., Bucher, P., Jongeneel, C.V., O'Hare,
 M.J., Soares, F., Brentani, R.K., Reis, L.F., de Souza, S.J. and
 Simpson, A.J.

TITLE Shotgun sequencing of the human transcriptome with ORF expressed
 sequence tags
 JOURNAL Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
 MEDLINE 20202663
 COMMENT

Contact: Simpson A.J.G.
 Laboratory of Cancer Genetics
 Ludwig Institute for Cancer Research
 Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
 Brazil

Tel: +55-11-2704922
 Fax: +55-11-2707001
 Email: asimpson@ludwig.org.br
 This sequence was derived from the FAPESP/LICR Human Cancer Genome
 Project. This entry can be seen in the following URL
 (http://www.ludwig.org.br/scripts/gethtml2.pl?tl=at2=RC4-HT0276-100
 300-015-f11&t3=2000-03-10&t4=1)
 Seq primer: puc 18 forward
 High quality sequence start: 2
 High quality sequence stop: 180.

FEATURES
 source
 1..180
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone_lib="HT0276"
 /dev_stage="Adult"
 /note="Organ: head_neck; Vector: puc18; Site_1: SmaI;
 Site_2: SmaI; A mini-library was made by cloning products
 derived from ORESTES PCR (U.S. Letters Patent application
 No. 196,716 - Ludwig Institute for Cancer Research)

profiles into the pUC 18 vector. Reverse transcription of
 tissue mRNA and cDNA amplification were performed under
 low stringency conditions."

BASE COUNT 31 a 47 c 51 g 51 t
 ORIGIN

Query Match 100.0%; Score 12; DB 10; Length 180;
 Best Local Similarity 100.0%; Pred. No. 3.6e+04;
 Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GTCGTGCAGGCA 12
 |||||

Db 47 GTCGTGCAGGCA 58

RESULT 5
 BI024277/c

LOCUS BI024277 192 bp mRNA linear EST 14-JUN-2001
 DEFINITION CM3-WT0293-260101-691-h12 MT0293 Homo sapiens cDNA, mRNA sequence.
 ACCESSION BI024277
 VERSION BI024277.1 GI:14430907
 KEYWORDS EST.
 SOURCE human.

ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Euthera; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 192)
 Dias Neto, E., Garcia Correa, R., Verjovski-Almeida, S., Briones, M.R.,
 Nagai, M.A., da Silva, W. Jr., Zago, M.A., Bordin, S., Costa, F.F.,
 Goldman, G.H., Carvalho, A.F., Matsukuma, A., Baia, G.S., Simpson, D.H.,
 Brunstein, A., deOliveira, P.S., Bucher, P., Jongeneel, C.V., O'Hare,
 M.J., Soares, F., Brentani, R.K., Reis, L.F., de Souza, S.J. and
 Simpson, A.J.

TITLE Shotgun sequencing of the human transcriptome with ORF expressed
 sequence tags
 JOURNAL Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
 MEDLINE 20202663
 COMMENT

Contact: Simpson A.J.G.
 Laboratory of Cancer Genetics
 Ludwig Institute for Cancer Research
 Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
 Brazil

Tel: +55-11-2704922
 Fax: +55-11-2707001
 Email: asimpson@ludwig.org.br
 This sequence was derived from the FAPESP/LICR Human Cancer Genome
 Project. This entry can be seen in the following URL
 (http://www.ludwig.org.br/scripts/gethtml2.pl?tl=CM3&t2=CM3-WT0293-
 260101-691-h12&t3=2001-01-26&t4=1)
 Seq primer: puc 18 forward
 High quality sequence start: 90
 High quality sequence stop: 190.

FEATURES
 source

1..192
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone_lib="MT0293"
 /dev_stage="Adult"
 /note="Organ: marrow; Vector: puc18; Site_1: SmaI; Site_2:
 SmaI; A mini-library was made by cloning products derived
 from ORESTES PCR (U.S. Letters Patent application No. 196
 716 - Ludwig Institute for Cancer Research) profiles
 into the pUC 18 vector. Reverse transcription of tissue
 mRNA and cDNA amplification were performed under low
 stringency conditions."

BASE COUNT 52 a 44 c 58 g 38 t
 ORIGIN

Query Match 100.0%; Score 12; DB 13; Length 192;
 Best Local Similarity 100.0%; Pred. No. 3.7e+04;
 Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GTCGTGCAGGCA 12

	Db	81	GTCTGTCAGGCCA 70		EST.	human. Homo sapiens Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo. REFERENCE PM2-NN0165-160301-007-e10 NN0165 Homo sapiens cDNA, mRNA sequence. BT034110 212 bp mRNA linear EST 14-JUN-2001 LOCUS BT034110.1 GI:14440736 DEFINITION ACCESSION VERSION KEYWORDS SOURCE ORGANISM Homo sapiens Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo. REFERENCE Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R., Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F., Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H., Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and Simpson,A.J. Shotgun sequencing of the human transcriptome with ORF expressed sequence tags Proc.Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000) 20202663 Contact: Simpson A.J.G. Laboratory of Cancer Genetics Ludwig Institute for Cancer Research Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP, Brazil Tel: +55-11-2704922 Fax: +55-11-2707001 Email: asimpson@ludwig.org.br This sequence was derived from the FAPESP/LICR Human Cancer Genome project. This entry can be seen in the following URL (http://www.ludwig.org.br/scripts/gethtml2.pl?tl=PM2&t2=PM2-NN0165- 160301-007-e10&t3=2001-03-16&t4=1) Seq primer: puc 18 forward High quality sequence stop: 212. Location/Qualifiers 1. .212 /organism="Homo sapiens" /db_xref="taxon:9606" /clone_lib="NN0165" /dev_stage="Adult" /note="Organ: nervous_normal; Vector: puc18; Site_1: SmaI; Site_2: SmaI; A mini-library was made by cloning products derived from ORESTES PCR (U.S. Letters Patent application No. 196,716 - Ludwig Institute for Cancer Research) profiles into the pUC 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions." BASE COUNT 50 a 72 c 50 g 40 t ORIGIN Query Match 100.0%; Score 12; DB 13; Length 212; Best Local Similarity 100.0%; Pred. No. 3.7e+04; Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0; QY 1 GTCTGTCAGGCCA 12 Db 172 GTCTGTCAGGCCA 161 RESULT 7 AA815339 LOCUS AA815339.1 DEFINITION ACCESSION VERSION KEYWORDS SOURCE ORGANISM Homo sapiens Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo. REFERENCE Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R., Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F., Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H., Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and Simpson,A.J. Shotgun sequencing of the human transcriptome with ORF expressed sequence tags Proc. Natl. Acad. Sci. U.S.A. 97 (7) 3491-3496 (2000)
	Db	81	GTCTGTCAGGCCA 70		EST.	human. Homo sapiens Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo. REFERENCE PM2-NN0165-160301-007-e10 NN0165 Homo sapiens cDNA, mRNA sequence. BT034110 212 bp mRNA linear EST 14-JUN-2001 LOCUS BT034110.1 GI:14440736 DEFINITION ACCESSION VERSION KEYWORDS SOURCE ORGANISM Homo sapiens Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo. REFERENCE Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R., Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F., Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H., Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and Simpson,A.J. Shotgun sequencing of the human transcriptome with ORF expressed sequence tags Proc.Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000) 20202663 Contact: Simpson A.J.G. Laboratory of Cancer Genetics Ludwig Institute for Cancer Research Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP, Brazil Tel: +55-11-2704922 Fax: +55-11-2707001 Email: asimpson@ludwig.org.br This sequence was derived from the FAPESP/LICR Human Cancer Genome project. This entry can be seen in the following URL (http://www.ludwig.org.br/scripts/gethtml2.pl?tl=PM2&t2=PM2-NN0165- 160301-007-e10&t3=2001-03-16&t4=1) Seq primer: puc 18 forward High quality sequence stop: 212. Location/Qualifiers 1. .212 /organism="Homo sapiens" /db_xref="taxon:9606" /clone_lib="NN0165" /dev_stage="Adult" /note="Organ: nervous_normal; Vector: puc18; Site_1: SmaI; Site_2: SmaI; A mini-library was made by cloning products derived from ORESTES PCR (U.S. Letters Patent application No. 196,716 - Ludwig Institute for Cancer Research) profiles into the pUC 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions." BASE COUNT 50 a 72 c 50 g 40 t ORIGIN Query Match 100.0%; Score 12; DB 13; Length 212; Best Local Similarity 100.0%; Pred. No. 3.7e+04; Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0; QY 1 GTCTGTCAGGCCA 12 Db 172 GTCTGTCAGGCCA 161 RESULT 7 AA815339 LOCUS AA815339.1 DEFINITION ACCESSION VERSION KEYWORDS SOURCE ORGANISM Homo sapiens Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo. REFERENCE Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R., Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F., Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H., Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and Simpson,A.J. Shotgun sequencing of the human transcriptome with ORF expressed sequence tags Proc. Natl. Acad. Sci. U.S.A. 97 (7) 3491-3496 (2000)

MEDLINE
COMMENT

20202663
Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Cancer Genome
project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml2.pl?tl=RC4&t2=RC4-CT0109-
311200-022-a03&t3=2000-12-31&t4=1)
311200-022-a03&t3=2000-12-31&t4=1)
Seq primer: puc 18 forward
High quality sequence start: 65
High quality sequence stop: 239.

FEATURES

Source

1..239
Location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="CT0109"
/dev_stage="Adult"

/note="Organ: colon; Vector: puc18; Site_1: SmaI; Site_2:
SmaI; A mini-library was made by cloning products derived
from ORESTES PCR (U.S. Letters Patent application No. 196
/716 - Ludwig Institute for Cancer Research) profiles
into the pUC 18 vector. Reverse transcription of tissue
mRNA and cDNA amplification were performed under low
stringency conditions."

BASE COUNT 58 a 68 c 55 g 58 t
ORIGIN

Query Match

Best Local Similarity 100.0%; Score 12; DB 12; Length 239;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GTCGTCGAGGCA 12
|||||
Db 42 GTCGTCGAGGCA 31

RESULT 9

BH227331/C
LOCUS BH227331 248 bp DNA linear GSS 08-NOV-2001
DEFINITION 1006139B07.y1 1006 - RescueMu Grid G Zea mays genomic, DNA
sequence.
ACCESSION BH227331
VERSION BH227331.1 GI:16827224
KEYWORDS GSS.
SOURCE Zea mays.
ORGANISM Zea mays

REFERENCE

AUTHORS Maize genomic sequences found using engineered RescueMu transposon
JOURNAL Unpublished (2001)
COMMENT Contact: Walbot V
Department of Biological Sciences
Stanford University
855 California Ave, Palo Alto, CA 94304, USA
Tel: 650 723 2227
Fax: 650 725 8221
Email: walbot@stanford.edu

Possible ligation site so sequence was trimmed. Post-ligation
sequence submitted separately.
Plate: 1006139 row: 4

Class: transposon-tagged.

FEATURES

Source

1..248
Location/Qualifiers
/organism="Zea mays"
/cultivar="mixed background W23/A188/B73"

/db_xref="taxon:4577"
/clone_lib="1006 - RescueMu Grid G"
/tissue_type="leaf"
/dev_stage="adult"
/lab_host="DH10B"
/note="Organ: leaf; Vector: RescueMu (engineered from
pBlueScript backbone); Site_1: BamHI; Site_2: BglII;
RescueMu is a 4.9 kb, modified maize Mu transposon
designed to allow plasmid rescue from total genomic DNA.
Mu elements insert preferentially into transcription
units. For more information on RescueMu, go to the web
site 'www.zmdb.tastate.edu' and follow the links for
'RescueMu.' Grid G was grown at Stanford in 2000. DNA was
extracted from leaf punches, double digested using BamHI
and BglII, and ligated to form circular plasmids. DH10B
cells were transformed and then screened on LB plates with
ampicillin." 40 a 80 c 89 g 39 t

Query Match 100.0%; Score 12; DB 17; Length 248;
Best Local Similarity 100.0%; Pred. No. 3.8e+04;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GTCGTCGAGGCA 12
|||||
Db 62 GTCGTCGAGGCA 51

RESULT 10

BH145815/C
LOCUS BH145815 253 bp mRNA linear EST 30-NOV-2001
DEFINITION TCAAPID7858 Pediatric acute myelogenous leukemia cell (FAB M1)
Baylor-HGSC project=TCAA Homo sapiens cDNA clone TCAAP7858, mRNA
sequence.
ACCESSION BH145815
VERSION BH145815.1 GI:17164180
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens

REFERENCE

AUTHORS Wei Y., Tsang Y.T.M., Mei G., Ku J.M., Ali-Osman, F.R. Jr.,
Gundaratne, P.H., Muzny, D., Bouck, J., Gibbs, R.A. and Margolin, J.F.
TITLE Pediatric Leukemia cDNA Sequencing Project (2001)
JOURNAL Unpublished (2001)
COMMENT Contact: Dr. Judith F. Margolin
Texas Children's Cancer Center and Human Genome Sequencing Center
at Baylor College of Medicine
1102 Bates, MC3-3320 Houston, TX 77030, USA
Tel: 832-824-4536
Fax: 832-825-4038
Email: clones@tccc.org

FEATURES

Source

1..253
Location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="TCAAP7858"
/clone_lib="Pediatric acute myelogenous leukemia cell (FAB
M1) Baylor-HGSC project=TCAA"

/tissue_type="leukopheresis"
/cell_type="myeloid cell"
/dev_stage="pediatric 6 years"
/lab_host="DH10B"

/note="Vector: lambda pSB; Site_1: BamHI; Site_2: EcoRI;
First strand cDNA was primed with an anchored
XhoI-oligo(dT) primer [5'GGAGGACTCGAGCGCCGAGGAG(T)VN
3'; V=A,C,G; N=A,C,G,T] and then dg tailed. Second strand
was primed with a BamHI-dC primer
[5'AGAGAGCTCGGATCGCGCGCCCAATAATAAT(C) 3'].

Mon Jan 6 15:20:20 2003

Double-stranded cDNA was then digested with BamHI and XhoI and directionally cloned into the BamHI and SalI sites of lambda PSB vector. Library was constructed by Wei Yu at RIKEN normalization. Library was constructed by Wei Yu at RIKEN of Japan (Carninci P, Westover A, Nishiyama Y, Ohsumi T, Itoh M, Nagaoka S, Sasaki N, Okazaki Y, Muramatsu M, Schneider C, Hayashizaki Y, High efficiency selection of full-length cDNA by improved biotinylated cap trapper., DNA Res 4: 1, 61-6, Feb 28, 1997)"

34 a 65 c 85 g 67 t 2 others

Query Match 100.0%; Score 12; DB 13; Length 253;
Best Local Similarity 100.0%; Pred. No. 3.8e+04;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GTCGTGCAGGCA 12

|||||

Db 218 GTCGTGCAGGCA 207

RESULT 11

BF256224/c

LOCUS

DEFINITION

BF256224 256 bp mRNA linear EST 22-OCT-2001
HVSMEf0009D01f Hordeum vulgare seedling root EST library HVCdNA0007
(Etolated and unstressed) Hordeum vulgare cDNA clone

HVSMEf0009D01f, mRNA sequence.

ACCESSION

BF256224

VERSION

BF256224.2

KEYWORDS

SOURCE

ORGANISM

Hordeum vulgare.

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae

; Triticeae; Hordeum.

1 (bases 1 to 256)

Wing, R., Close, T.J., Klein, H.A., Wise, R., Begum, D., Frisch, D., Yu

Y., Henry, D., Palmer, M., Rambo, T., Simmons, J., Choi, D.W., Fenton

R.D., Oates, R. and Main, D.

Development of a genetically and physically anchored EST resource

for barley genomics: Morex unstressed seedling root cDNA library

Unpublished (2001)

On Nov 16, 2000 this sequence version replaced gi:11185337.

Contact: Wing RA

Clemson University

Genomics Institute

Clemson University

100 Jordan Hall, Clemson, SC 29634, USA

Tel: 864 656 7288

Fax: 864 656 4293

Email: rwing@clemson.edu

Total hg bases = 104

Seq primer: AATTAACCTCCTAAAGG

High quality sequence start: 28

High quality sequence stop: 255.

Location/Qualifiers

1. .256

/organism="Hordeum vulgare"

/cultivar="Morex"

/db_xref="taxon:4513"

/clone="HVSMEf0009D01f"

HVCdNA0007 (Etolated and unstressed)

/tissue_type="Seedling root"

/lab_host="TJC121"

/note="Vector: lambdaZAP; Site 1: EcoRI; Site 2: XhoI;

Seeds were surface sterilized then germinated under axenic

conditions in the dark at room temperature on filter paper

with water, nystatin and cefotaxime in covered

crystallization dishes. Five-day old seedling roots were

then harvested, total RNA was prepared, poly(A) RNA was

purified, one primary unamplified cDNA library was made,

and 1 million pfu were in vivo excised to give pBluescript

SK(-) cDNA phagemids. These steps were performed in the TJ

Close laboratory at the University of California, Riverside (Choi, Close, Fenton). Phagemids were plated and picked at the Clemson University Genomics Institute (CUGI) (Begum, Palmer, Frisch, Atkins and Wing). Plasmid DNA preparations, DNA sequencing and sequence analysis were performed at CUGI (Wing, Yu, Frisch, Henry, Simmons, Oates, Rambo, Main). The sequence has been trimmed to remove vector sequence and contains a minimum of 100 bases of phred value 20 or above. For more details on library preparation and sequence analysis see <http://www.genome.clemson.edu/projects/barley>. To order this clone see <http://www.genome.clemson.edu/orders> Also see Close TJ, Wing R, Kleinhofs A, Wise R (2001) Genetically and physically anchored EST resources for barley genomics. Barley Genetics Newsletter 31:29-30. (<http://wheat.pw.usda.gov/ggpages/bgn/31/cover.html>)"

BASE COUNT 57 a 84 c 73 g 42 t

ORIGIN

Query Match 100.0%; Score 12; DB 12; Length 256;

Best Local Similarity 100.0%; Pred. No. 3.8e+04;

Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GTCGTGCAGGCA 12

|||||

Db 221 GTCGTGCAGGCA 210

RESULT 12

BF146813/c

LOCUS

DEFINITION

BF146813 261 bp mRNA linear EST 29-DEC-2000
uy36f05.x1 NCI_CGAP_Lu30 Mus musculus cDNA clone IMAGE:3661665 3'
Similar to SW:SMB2_MOUSE P40694 DNA-BINDING PROTEIN SMUBP-2 ; mRNA
sequence.

ACCESSION

BF146813

KEYWORDS

EST.

SOURCE

ORGANISM

Mus musculus

house mouse.

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

1 (bases 1 to 261)

NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.

National Cancer Institute, Cancer Genome Anatomy Project (CGAP),

Tumor Gene Index

Unpublished (1997)

Contact: Robert Strausberg, Ph.D.

Email: cgapbs-r@mail.nih.gov

Tissue Procurement: Gilbert Smith, Ph.D.

cDNA Library Preparation: Life Technologies, Inc.

DNA Sequencing by: The I.M.A.G.E. Consortium (LLNL)

Clone distribution: NCI-CGAP clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:

image.llnl.gov/image/html/iresources.shtml

MG1:1422433

Possible reversed clone: polyT not found

Seq primer: -400P from Gibco.

Location/Qualifiers

1. .261

/organism="Mus musculus"

/strain="CZECH II"

/db_xref="taxon:10090"

/clone="IMAGE:3661665"

/clone_lib="NCI_CGAP_Lu30"

/tissue_type="tumor, metastatic to mammary"

/lab_host="DH10B"

/note="Organ: lung; Vector: pCMV-SPORT6; Site 1: NotI;

Site 2: SalI; transgenic model WNT-1, expression driven by

MMTV-LTR enhancer; Cloned unidirectionally. Primer: Oligo

dT. Library constructed by Life Technologies.

Investigator providing samples: Gilbert Smith, NIH"

FEATURES

source

Mon Jan 6 15:20:20 2003

priming with dt-tailed vector. The dt-tailed vector was adjusted to have about 60nt. The cDNA vector was circularized with E. coli DNA ligase after digestion of EcoRI which site is also included in vector. An RNA strand converted to a DNA strand by Okayama-Berg method. The obtained cDNA vectors were used for transformation of competent cells E. coli Top10F⁺ by electroporation method. The cDNA libraries constructed by this method are full-length enriched cDNA library."

BASE COUNT 39 a 96 c 80 g 77 t
 ORIGIN
 Query Match 100.0%; Score 12; DB 14; Length 292;
 Best Local Similarity 100.0%; Pred. No. 3.9e+04;
 Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GTCGTCGAGCA 12
 |||||
 Db 32 GTCGTCGAGCA 43

RESULT 15
 AT004345/C
 LOCUS 296 bp mRNA linear EST 25-MAR-2002
 DEFINITION AT004345 POSLM01 Pleurotus ostreatus cDNA clone 1893LM, mRNA
 sequence.
 ACCESSION AT004345
 VERSION AT004345.1 GI:13419203
 KEYWORDS EST.
 SOURCE oyster mushroom.
 ORGANISM Pleurotus ostreatus
 Eukaryota; Fungi; Basidiomycota; Hymenomycetes; Homobasidiomycetes;
 Agaricales; Pleurotaceae; Pleurotus.
 REFERENCE 1 (bases 1 to 296)
 AUTHORS Lee, S.H., Kim, B.G., Kim, K.J., Lee, J.S., Yun, D.W., Hahn, J.H., Kim, G.H., Lee, K.H., Suh, D.S., Kwon, S.T., Lee, C.S. and Yoo, Y.B.
 TITLE Comparative Analysis of Sequences Expressed during the Liquid-Cultured Mycelia and Fruit Body Stages of Pleurotus ostreatus
 JOURNAL Fungal Genet. Biol. 35 (2), 115-134 (2002)
 MEDLINE 21838665
 COMMENT Contact: Beom-Gi Kim
 Division of applied microbiology
 Institute of Agricultural Science and Technology(NIAST)
 249 Seodundong Kwonseonku, Suwon 441707, Korea
 Tel: 82-331-290-0347
 Fax: 82-331-290-0399
 Email: bckimyes@da.go.kr
 GeneNuri No. KSL04602
 Submitted through BRIC(Biological Research Information Center) of Korea
 URL: http://bric.postech.ac.kr/.

FEATURES
 source
 1. .296
 Location/Qualifiers
 /organism="Pleurotus ostreatus"
 /cultivar="AST 2029"
 /db_xref="taxon:5322"
 /clone="1893LM"
 /clone_lib="PSLM01"
 /dev_stage="shaking liquid cultured mycelia"
 /lab_host="E.coli"
 /note="vector: lambda Uni-ZAP XR; Site.1: EcoRI; Site.2: XhoI; average insert size:1000 bp;initial pfu:5 x 10⁷
 Library information:Isolation of total RNA from the mycelia incubated in shaking liquid MCM media at 30 deg C"
 BASE COUNT 97 a 80 c 60 g 57 t 2 others
 ORIGIN
 Query Match 100.0%; Score 12; DB 9; Length 296;
 Best Local Similarity 100.0%; Pred. No. 3.9e+04;
 Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GTCGTCGAGCA 12

Db 273 GTCGTCGAGCA 262
 |||||

Search completed: January 4, 2003, 01:04:23
 Job time : 164.558 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 3, 2003, 22:53:16 ; Search time 4.05047 Seconds
(without alignments)
908.566 Million cell updates/sec

Title: US-09-787-562-10
Perfect score: 12
Sequence: 1 gtcgtgcaggca 12

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 441362 seqs, 15338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NA.*
1: /cgn2_6/ptodata/2/ina/5A_COMB.seq.*
2: /cgn2_6/ptodata/2/ina/5B_COMB.seq.*
3: /cgn2_6/ptodata/2/ina/6A_COMB.seq.*
4: /cgn2_6/ptodata/2/ina/6B_COMB.seq.*
5: /cgn2_6/ptodata/2/ina/PTUS_COMB.seq.*
6: /cgn2_6/ptodata/2/ina/backfiles1.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	12	100.0	35	1	US-08-360-841-4
C 2	12	100.0	36	1	US-08-360-841-11
C 3	12	100.0	174	1	US-08-360-841-1
C 4	12	100.0	953	4	US-08-438-745-3
C 5	12	100.0	953	4	US-08-438-745-5
C 6	12	100.0	953	4	US-09-219-019-3
C 7	12	100.0	953	4	US-09-219-019-5
C 8	12	100.0	953	5	PCT-US94-05669A-3
C 9	12	100.0	953	5	PCT-US94-05669A-5
C 10	12	100.0	2087	4	US-09-097-199-83
C 11	12	100.0	7859	1	US-07-854-596B-4
C 12	12	100.0	7859	2	US-08-450-905B-15
C 13	12	100.0	7859	3	US-07-982-759F-15
C 14	11	91.7	30	4	US-09-327-984A-13
C 15	11	91.7	30	4	US-09-327-984A-14
C 16	11	91.7	41	4	US-09-327-984A-28
C 17	11	91.7	41	4	US-09-327-984A-15
C 18	11	91.7	49	4	US-09-327-984A-16
C 19	11	91.7	49	4	US-09-327-984A-17
C 20	11	91.7	49	4	US-09-327-984A-18
C 21	11	91.7	49	4	US-09-327-984A-19
C 22	11	91.7	282	2	US-08-105-989-10
C 23	11	91.7	282	3	US-09-138-922-10
C 24	11	91.7	416	4	US-09-319-056B-1
C 25	11	91.7	416	4	US-09-319-056B-3
C 26	11	91.7	614	2	US-08-729-103-2
C 27	11	91.7	652	4	US-08-998-416-962

28	11	91.7	861	4	US-08-998-416-299	Sequence 299, Appl
C 29	11	91.7	957	2	US-08-544-822-2	Sequence 2, Appli
C 30	11	91.7	957	3	US-09-070-964-2	Sequence 2, Appli
31	11	91.7	1060	3	US-09-188-930-9	Sequence 9, Appli
32	11	91.7	1080	4	US-09-125-642C-3	Sequence 3, Appli
C 33	11	91.7	1114	2	US-08-468-413-1	Sequence 1, Appli
C 34	11	91.7	1114	3	US-09-162-508-1	Sequence 1, Appli
C 35	11	91.7	1114	5	PCT-US95-07169-1	Sequence 1, Appli
C 36	11	91.7	1193	1	US-07-956-697B-4	Sequence 4, Appli
C 37	11	91.7	1193	1	US-08-263-098-4	Sequence 4, Appli
38	11	91.7	1193	4	US-09-541-941B-27	Sequence 27, Appl
C 39	11	91.7	1246	1	US-08-446-777-3	Sequence 3, Appli
C 40	11	91.7	1249	4	US-09-333-208-1	Sequence 1, Appli
C 41	11	91.7	1249	4	US-09-333-254-1	Sequence 1, Appli
C 42	11	91.7	1249	4	US-09-183-270-1	Sequence 1, Appli
43	11	91.7	1280	3	US-09-188-930-246	Sequence 246, App
44	11	91.7	1445	1	US-08-324-533-1	Sequence 1, Appli
45	11	91.7	1501	3	US-08-993-359-21	Sequence 21, Appl

ALIGNMENTS

RESULT 1
US-08-360-841-4/c
; Sequence 4, Application US/08360841
; Patent No. 5652120
; GENERAL INFORMATION:
; APPLICANT: PARK, Seung Kook
; APPLICANT: LEE, Kang Moon
; APPLICANT: NHO, Kyoo Seung
; APPLICANT: KOH, Yeo Wook
; APPLICANT: KWON, Chang Hyuk
; APPLICANT: CHUNG, Ju Young
; APPLICANT: JEE, Young Su
; APPLICANT: YU, Young Hyo
; TITLE OF INVENTION: A No. 5652120e1 Gene Coding Human Epidermal
; TITLE OF INVENTION: Growth Factor and Process for Preparing the Same
; NUMBER OF SEQUENCES: 20
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Darby & Darby PC
; STREET: 805 Third Avenue
; CITY: New York
; STATE: New York
; COUNTRY: US
; ZIP: 10022
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/360,841
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Ludwig, S. Peter
; REGISTRATION NUMBER: 25,351
; REFERENCE/DOCKET NUMBER: 0136/0A760
; TELEPHONE: 212-527-7700
; TELEFAX: 212-753-6237
; TELEX: 236687
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 35 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA to mRNA
; ORIGINAL SOURCE:
; ORGANISM: Homo sapiens
; IMMEDIATE SOURCE:

CLONE: C2 PRIMER
US-08-360-841-4

Query Match 100.0%; Score 12; DB 1; Length 35;
Best Local Similarity 100.0%; Pred. NO. 2.3e+02;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GTCGTGCAGGCA 12
|||||
Db 24 GTCGTGCAGGCA 13

RESULT 2

US-08-360-841-11
; Sequence 11, Application US/08360841

; Patent No. 5652120

; GENERAL INFORMATION:

; APPLICANT: PARK, Seung Kook

; APPLICANT: LEE, Kang Moon

; APPLICANT: NHO, Kyo Seung

; APPLICANT: KOH, Yeo Wook

; APPLICANT: KWON, Chang Hyuk

; APPLICANT: CHUNG, Ju Young

; APPLICANT: JEE, Young Su

; APPLICANT: YU, Young Hyo

; TITLE OF INVENTION: A No. 5652120el Gene Coding Human Epidermal

; GROWTH FACTOR AND PROCESS FOR PREPARING THE SAME

; NUMBER OF SEQUENCES: 20

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Darby & Darby PC

; STREET: 805 Third Avenue

; CITY: New York

; STATE: New York

; COUNTRY: US

; ZIP: 10022

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/360,841

; FILING DATE:

; CLASSIFICATION: 435

; ATTORNEY/AGENT INFORMATION:

; NAME: Ludwig, S. Peter

; REGISTRATION NUMBER: 25,351

; REFERENCE/DOCKET NUMBER: 0136/OA760

; TELEPHONE: 212-527-7700

; TELEFAX: 212-753-6237

; TELEX: 236687

; INFORMATION FOR SEQ ID NO: 11:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 36 base pairs

; TYPE: nucleic acid

; STRANDEDNESS: single

; TOPOLOGY: linear

; MOLECULE TYPE: cDNA to mRNA

; ORGANISM: Homo sapiens

; IMMEDIATE SOURCE:

; CLONE: N4 PRIMER

US-08-360-841-11

Query Match 100.0%; Score 12; DB 1; Length 36;
Best Local Similarity 100.0%; Pred. NO. 2.3e+02;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GTCGTGCAGGCA 12
|||||
Db 21 GTCGTGCAGGCA 32

RESULT 3

US-08-360-841-1/c

; Sequence 1, Application US/08360841

; Patent No. 5652120

; GENERAL INFORMATION:

; APPLICANT: PARK, Seung Kook

; APPLICANT: LEE, Kang Moon

; APPLICANT: NHO, Kyo Seung

; APPLICANT: KOH, Yeo Wook

; APPLICANT: KWON, Chang Hyuk

; APPLICANT: CHUNG, Ju Young

; APPLICANT: JEE, Young Su

; APPLICANT: YU, Young Hyo

; TITLE OF INVENTION: A No. 5652120el Gene Coding Human Epidermal

; GROWTH FACTOR AND PROCESS FOR PREPARING THE SAME

; NUMBER OF SEQUENCES: 20

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Darby & Darby PC

; STREET: 805 Third Avenue

; CITY: New York

; STATE: New York

; COUNTRY: US

; ZIP: 10022

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/360,841

; FILING DATE:

; CLASSIFICATION: 435

; ATTORNEY/AGENT INFORMATION:

; NAME: Ludwig, S. Peter

; REGISTRATION NUMBER: 25,351

; REFERENCE/DOCKET NUMBER: 0136/OA760

; TELEPHONE: 212-527-7700

; TELEFAX: 212-753-6237

; TELEX: 236687

; INFORMATION FOR SEQ ID NO: 1:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 174 base pairs

; TYPE: nucleic acid

; STRANDEDNESS: single

; TOPOLOGY: linear

; MOLECULE TYPE: cDNA to mRNA

; ORGANISM: Homo sapiens

; IMMEDIATE SOURCE:

; CLONE: EGF-nt seq

US-08-360-841-1

Query Match 100.0%; Score 12; DB 1; Length 174;

Best Local Similarity 100.0%; Pred. NO. 2.2e+02;

Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GTCGTGCAGGCA 12

|||||

Db 54 GTCGTGCAGGCA 43

RESULT 4

US-08-438-745-3/c

; Sequence 3, Application US/08438745

; Patent No. 6248715

; GENERAL INFORMATION:

; APPLICANT: Rosenberg, Steven

; APPLICANT: Stratton-Thomas, Jennifer

; TITLE OF INVENTION: Expression of Urokinase Plasminogen

; ACTIVATOR INHIBITORS

; NUMBER OF SEQUENCES: 22


```

;
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Chiron Corporation
; STREET: 4560 Horton Street
; CITY: Emeryville
; STATE: CA
; COUNTRY: USA
; ZIP: 94608
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
;
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/438,745
; FILING DATE: 10-MAY-1995
; CLASSIFICATION: 514
;
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/070,153
; FILING DATE: 01-JUN-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Green, Grant D.
; REGISTRATION NUMBER: 31,259
; REFERENCE/DOCKET NUMBER: 0939.001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 510-601-2706
; TELEFAX: 510-655-3542
;
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 953 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: CDNA
; HYPOTHETICAL: NO
; IMMEDIATE SOURCE:
; CLONE: M1Flag-EGF-pIII fusion
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 25..903
;
; US-08-438-745-3
;
; Query Match 100.0%; Score 12; DB 4; Length 953;
; Best Local Similarity 100.0%; Pred. No. 2.2e+02;
; Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
;
; QY 1 GTCGTGCAGGCA 12
; |||||
; DB 165 GTCGTGCAGGCA 154
;
; RESULT 5
; US-08-438-745-5/c
; Sequence 5, Application US/08438745
; Patent No. 6248715
; GENERAL INFORMATION:
; APPLICANT: Rosenberg, Steven
; TITLE OF INVENTION: Expression of Urokinase Plasminogen
; TITLE OF INVENTION: Activator Inhibitors
; NUMBER OF SEQUENCES: 22
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Chiron Corporation
; STREET: 4560 Horton Street
; CITY: Emeryville
; STATE: CA
; COUNTRY: USA
; ZIP: 94608
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
;
; CURRENT APPLICATION DATA:

```

```

;
; APPLICATION NUMBER: US/08/438,745
; FILING DATE: 10-MAY-1995
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/070,153
; FILING DATE: 01-JUN-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Green, Grant D.
; REGISTRATION NUMBER: 31,259
; REFERENCE/DOCKET NUMBER: 0939.001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 510-601-2706
; TELEFAX: 510-655-3542
;
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 953 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: CDNA
; HYPOTHETICAL: NO
; IMMEDIATE SOURCE:
; CLONE: M1Flag-EGF-pIII fusion
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 25..903
;
; US-08-438-745-5

```

```

Query Match 100.0%; Score 12; DB 4; Length 953;
Best Local Similarity 100.0%; Pred. No. 2.2e+02;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 1 GTCGTGCAGGCA 12
|||
DB 165 GTCGTGCAGGCA 154

```

```

RESULT 6
US-09-219-019-3/c
; Sequence 3, Application US/09219019
; Patent No. 6268341
; GENERAL INFORMATION:
; APPLICANT: ROSENBERG, STEVEN
; APPLICANT: STRATTON-THOMAS, JENNIFER R.
; TITLE OF INVENTION: EXPRESSION OF UROKINASE PLASMINOGEN ACTIVATOR
; TITLE OF INVENTION: INHIBITORS
; FILE REFERENCE: 23533-0005
; CURRENT APPLICATION NUMBER: US/09/219,019
; CURRENT FILING DATE: 1998-12-23
; PRIOR APPLICATION NUMBER: 08/438,263
; PRIOR FILING DATE: 1995-05-10
; PRIOR APPLICATION NUMBER: 08/280,288
; PRIOR FILING DATE: 1994-07-26
; PRIOR APPLICATION NUMBER: 08/070,153
; PRIOR FILING DATE: 1993-06-01
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3
; LENGTH: 953
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (25)..(903)
;
; US-09-219-019-3

```

```

Query Match 100.0%; Score 12; DB 4; Length 953;
Best Local Similarity 100.0%; Pred. No. 2.2e+02;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 1 GTCGTGCAGGCA 12
|||
DB 165 GTCGTGCAGGCA 154

```

LENGTH: 953 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: cdna
HYPOTHETICAL: NO
IMMEDIATE SOURCE:
CLONE: M1flag-EGF-pIII fusion
FEATURE:
NAME/KEY: CDS
LOCATION: 25..903
PCT-US94-05669A-3

Query Match 100.0%; Score 12; DB 5; Length 953;
Best Local Similarity 100.0%; Pred. No. 2.2e+02;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GTCGTGCAGGCA 12
|||||
Db 165 GTCGTGCAGGCA 154

RESULT 9
PCT-US94-05669A-5/c
Sequence 5, Application PC/TUS9405669A
GENERAL INFORMATION:
APPLICANT: Chiron Corporation
TITLE OF INVENTION: Expression of Urokinase Plasminogen
ACTIVATOR INHIBITORS
NUMBER OF SEQUENCES: 22
CORRESPONDENCE ADDRESS:
ADDRESSEE: Chiron Corporation
STREET: 4560 Horton Street
CITY: Emeryville
STATE: CA
COUNTRY: USA
ZIP: 94608
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US94/05669A
FILING DATE: 19-MAY-1994
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Green, Grant D.
REGISTRATION NUMBER: 31,259
REFERENCE/DOCKET NUMBER: 0939,100
TELECOMMUNICATION INFORMATION:
TELEPHONE: 510-601-2706
TELEFAX: 510-655-3542
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 953 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: cdna
HYPOTHETICAL: NO
IMMEDIATE SOURCE:
CLONE: M1flag-EGF-pIII fusion
FEATURE:
NAME/KEY: CDS
LOCATION: 25..903
PCT-US94-05669A-5

Query Match 100.0%; Score 12; DB 5; Length 953;
Best Local Similarity 100.0%; Pred. No. 2.2e+02;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GTCGTGCAGGCA 12
|||||
Db 165 GTCGTGCAGGCA 154

LENGTH: 953
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: CDS
LOCATION: (25)..(903)
US-09-219-019-5

Query Match 100.0%; Score 12; DB 4; Length 953;
Best Local Similarity 100.0%; Pred. No. 2.2e+02;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GTCGTGCAGGCA 12
|||||
Db 165 GTCGTGCAGGCA 154

RESULT 8
PCT-US94-05669A-3/c
Sequence 3, Application PC/TUS9405669A
GENERAL INFORMATION:
APPLICANT: Chiron Corporation
TITLE OF INVENTION: Expression of Urokinase Plasminogen
ACTIVATOR INHIBITORS
NUMBER OF SEQUENCES: 22
CORRESPONDENCE ADDRESS:
ADDRESSEE: Chiron Corporation
STREET: 4560 Horton Street
CITY: Emeryville
STATE: CA
COUNTRY: USA
ZIP: 94608
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US94/05669A
FILING DATE: 19-MAY-1994
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Green, Grant D.
REGISTRATION NUMBER: 31,259
REFERENCE/DOCKET NUMBER: 0939,100
TELECOMMUNICATION INFORMATION:
TELEPHONE: 510-601-2706
TELEFAX: 510-655-3542
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:

Db 165 GTCGTGAGCA 154
|||||

RESULT 10

US-09-097-199-83/c
; Sequence 83, Application US/09097199
; Patent No. 6218529
; GENERAL INFORMATION:
; APPLICANT: An, Gang
; APPLICANT: O'Hara, S. Mark
; APPLICANT: Ralph, David
; APPLICANT: Velti, Robert
; TITLE OF INVENTION: BIOMARKERS AND TARGETS FOR DIAGNOSIS,
; PROGNOSIS AND MANAGEMENT OF PROSTATE DISEASE
; NUMBER OF SEQUENCES: 87
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Arnold, White & Durkee
; STREET: P.O. Box 4433
; CITY: Houston
; STATE: Texas
; COUNTRY: USA
; ZIP: 77210
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/097,199
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/692,787
; FILING DATE: 31-JUL-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Nakashima, Richard A.
; REGISTRATION NUMBER: P-42,023
; REFERENCE/DOCKET NUMBER: UROC:018
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (512) 418-3000
; TELEFAX: (512) 474-7577
; INFORMATION FOR SEQ ID NO: 83:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2087 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 99..503
; US-09-097-199-83

Query Match 100.0%; Score 12; DB 4; Length 2087;
Best Local Similarity 100.0%; Pred. No. 2.2e+02;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GTCGTGAGCA 12
|||||

Db 1967 GTCGTGAGCA 1956

RESULT 11

US-07-854-596B-4/c
; Sequence 4, Application US/07854596B
; Patent No. 5434073
; GENERAL INFORMATION:
; APPLICANT: Dawson, Keith M
; APPLICANT: Hunter, Michael G
; APPLICANT: Czaplewski, Lloyd G
; TITLE OF INVENTION: Proteins and nucleic acids
; NUMBER OF SEQUENCES: 73
; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Dr. John J. McDonnell
; STREET: Ten South Wacker Drive, Suite 3000
; CITY: Chicago
; STATE: IL
; COUNTRY: USA
; ZIP: 60606
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/854,596B
; FILING DATE: 03-JUN-1992
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: McDonnell, John J.
; REGISTRATION NUMBER: 26,949
; REFERENCE/DOCKET NUMBER: 92,337
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 312-715-1000
; TELEFAX: 312-715-1234
; TELEX: 910-221-5317
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 7859 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: circular
; MOLECULE TYPE: CDNA
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: 1..7859
; OTHER INFORMATION: /note= "sequence of plasmid psw6"
; US-07-854-596B-4

Query Match 100.0%; Score 12; DB 1; Length 7859;
Best Local Similarity 100.0%; Pred. No. 2.2e+02;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GTCGTGAGCA 12
|||||

Db 7739 GTCGTGAGCA 7728

RESULT 12

US-08-450-905B-15/c
; Sequence 15, Application US/08450905B
; Patent No. 5856301
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: Stem Cell Inhibiting Proteins
; NUMBER OF SEQUENCES: 178
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: HALE and DORR
; STREET: 60 State Street
; CITY: Boston
; STATE: MA
; ZIP: 02109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/450,905B
; FILING DATE: 26-MAR-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/982,759
; FILING DATE: 08-MAR-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GB 9127319.3
; FILING DATE: 23-DEC-1991

```
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GB 9221587.0
; FILING DATE: 14-OCT-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: BAKER, HOLLIE L.
; REGISTRATION NUMBER: 31,321
; REFERENCE/DOCKET NUMBER: 102.378.120DV-2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617-526-6110
; TELEFAX: 617-526-5000
; INFORMATION FOR SEQ ID NO: 15:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 7859 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: circular
; MOLECULE TYPE: DNA
US-08-450-905B-15

Query Match 100.0%; Score 12; DB 2; Length 7859;
Best Local Similarity 100.0%; Pred. No. 2.2e+02;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GTCGTGCAGGCA 12
|||||
Db 7739 GTCGTGCAGGCA 7728

RESULT 13
US-07-982-759F-15/c
; Sequence 15, Application US/07982759F
; Patent No. 6057123
; GENERAL INFORMATION:
; APPLICANT: CRAIG, Stewart
; APPLICANT: GEORGE, Michael
; APPLICANT: EDWARDS, Richard Mark
; APPLICANT: CZAPLEWSKI, Lloyd George
; APPLICANT: GILBERT, Richard
; TITLE OF INVENTION: Stem Cell Inhibiting Proteins
; NUMBER OF SEQUENCES: 178
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: HALE and DORR LLP
; STREET: 60 State Street
; CITY: Boston
; STATE: MA
; ZIP: 02109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/982,759F
; FILING DATE: 08-MAR-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GB 9127319.3
; FILING DATE: 23-DEC-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GB 9221587.0
; FILING DATE: 14-OCT-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: BAKER, HOLLIE L.
; REGISTRATION NUMBER: 31,321
; REFERENCE/DOCKET NUMBER: 102378.120
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617-526-6000
; TELEFAX: 617-526-5000
; INFORMATION FOR SEQ ID NO: 15:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 7859 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: circular
```

```
; MOLECULE TYPE: DNA
US-07-982-759F-15

Query Match 100.0%; Score 12; DB 3; Length 7859;
Best Local Similarity 100.0%; Pred. No. 2.2e+02;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GTCGTGCAGGCA 12
|||||
Db 7739 GTCGTGCAGGCA 7728

RESULT 14
US-09-327-984A-13/c
; Sequence 13, Application US/09327984A
; Patent No. 6368594
; GENERAL INFORMATION:
; APPLICANT: Doetsch, Paul W.
; APPLICANT: Kaur, Balveen
; APPLICANT: Avery, Angela M.
; TITLE OF INVENTION: Broad Specificity DNA Damage Endonuclease
; FILE REFERENCE: 25-98
; CURRENT APPLICATION NUMBER: US/09/327,984A
; CURRENT FILING DATE: 1999-06-08
; PRIOR APPLICATION NUMBER: US 60/088,521
; PRIOR FILING DATE: 1998-06-08
; PRIOR APPLICATION NUMBER: US 60/134,752
; PRIOR FILING DATE: 1999-05-18
; NUMBER OF SEQ ID NOS: 39
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 13
; LENGTH: 30
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:double stranded
; OTHER INFORMATION: oligonucleotide containing cis-syn cyclobutane
; OTHER INFORMATION: pyrimidine dimer
; NAME/KEY: misc.feature
; LOCATION: (15)..(16)
; OTHER INFORMATION: At positions 15-16, the T-T is in the form of a
; OTHER INFORMATION: cis-syn cyclobutane pyrimidine dimer
US-09-327-984A-13

Query Match 91.7%; Score 11; DB 4; Length 30;
Best Local Similarity 100.0%; Pred. No. 8.5e+02;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

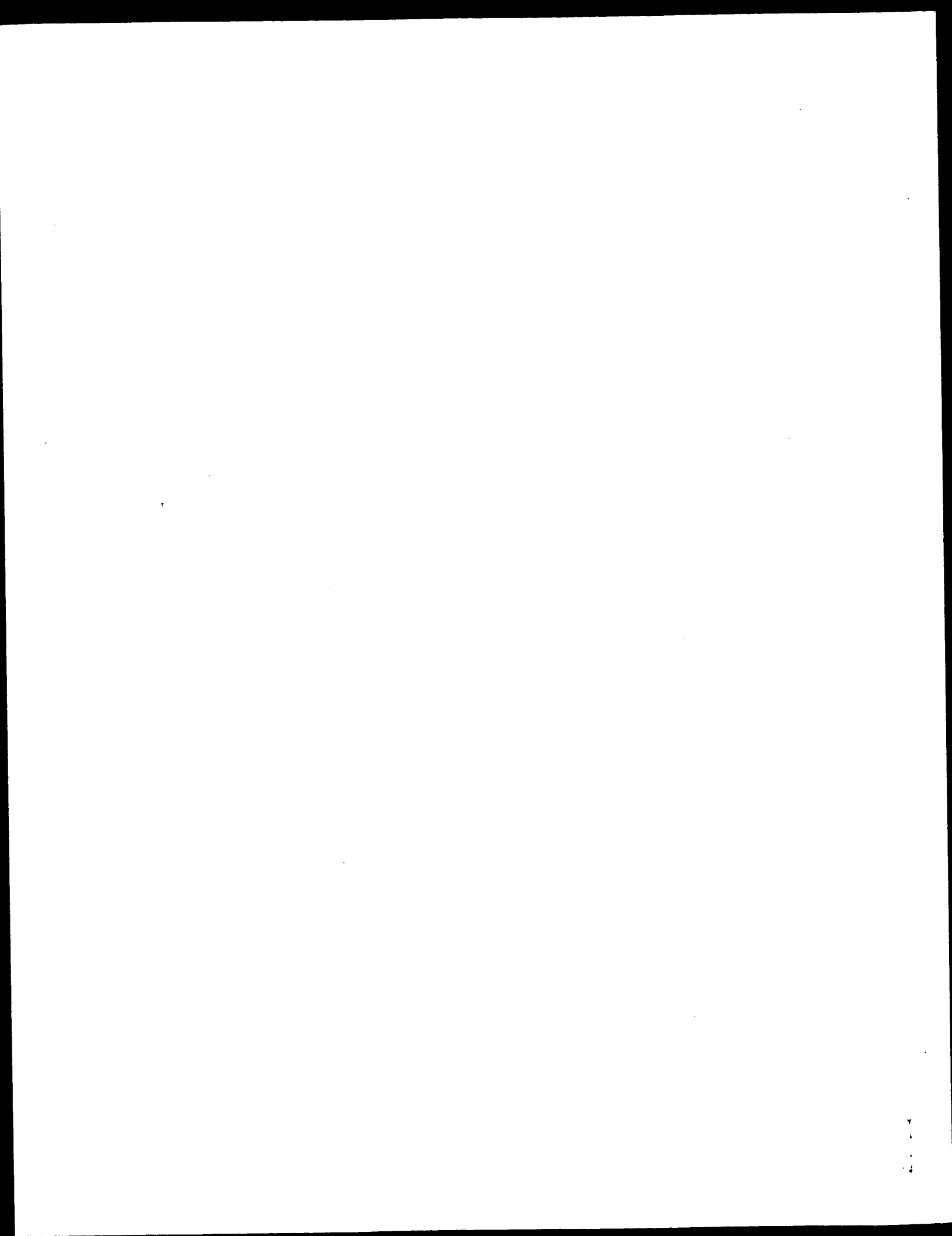
QY 2 TCGTGCAGGCA 12
|||||
Db 13 TCGTGCAGGCA 3

RESULT 15
US-09-327-984A-14/c
; Sequence 14, Application US/09327984A
; Patent No. 6368594
; GENERAL INFORMATION:
; APPLICANT: Doetsch, Paul W.
; APPLICANT: Kaur, Balveen
; APPLICANT: Avery, Angela M.
; TITLE OF INVENTION: Broad Specificity DNA Damage Endonuclease
; FILE REFERENCE: 25-98
; CURRENT APPLICATION NUMBER: US/09/327,984A
; CURRENT FILING DATE: 1999-06-08
; PRIOR APPLICATION NUMBER: US 60/088,521
; PRIOR FILING DATE: 1998-06-08
; PRIOR APPLICATION NUMBER: US 60/134,752
; PRIOR FILING DATE: 1999-05-18
; NUMBER OF SEQ ID NOS: 39
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 14
; LENGTH: 30
```

; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:undamaged
; OTHER INFORMATION: double stranded oligonucleotide
US-09-327-984A-14

Query Match 91.7%; Score 11; DB 4; Length 30;
Best Local Similarity 100.0%; Pred. No. 8.5e+02;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 2 TCGTGCAGGCA 12
|||||
Db 13 TCGTGCAGGCA 3

Search completed: January 4, 2003, 00:10:07
Job time : 6.05047 secs



GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 3, 2003, 22:54:41 ; Search time 4.05047 Seconds
(without alignments)
1281.345 Million cell updates/sec

Title: US-09-787-562-10

Perfect score: 12

Sequence: 1 gtcgtcaggca 12

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 381593 seqs, 216252194 residues

Total number of hits satisfying chosen parameters: 763186

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- Published Applications, NA: *
- 1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:*
 - 2: /cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq:*
 - 3: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq:*
 - 4: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq:*
 - 5: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq:*
 - 6: /cgn2_6/ptodata/1/pubpna/PCTUS_PUBCOMB.seq:*
 - 7: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq:*
 - 8: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:*
 - 9: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq:*
 - 10: /cgn2_6/ptodata/1/pubpna/US09_PUBCOMB.seq:*
 - 11: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq:*
 - 12: /cgn2_6/ptodata/1/pubpna/US10_PUBCOMB.seq:*
 - 13: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq:*
 - 14: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	12	100.0	355	10	US-09-560-863-114
2	12	100.0	575	10	US-09-917-800A-992
3	12	100.0	1351	10	US-09-731-872-80
4	12	100.0	1474	10	US-09-731-872-139
5	12	100.0	1677	10	US-09-915-582-16
6	12	100.0	1802	10	US-09-799-777-104
7	12	100.0	1845	9	US-09-738-626-3487
8	12	100.0	1848	10	US-09-915-582-34
9	12	100.0	11009	10	US-09-845-583-1
10	11	91.7	118	10	US-09-867-703-6217
11	11	91.7	126	10	US-09-815-343-1511
12	11	91.7	200	9	US-09-900-714A-3
13	11	91.7	247	10	US-09-923-876-557
14	11	91.7	274	10	US-09-960-352-10276
15	11	91.7	286	9	US-09-841-157A-2
16	11	91.7	286	10	US-09-294-093B-2281
17	11	91.7	287	10	US-09-294-093B-3949
18	11	91.7	300	10	US-09-294-093B-3480
19	11	91.7	326	10	US-09-764-864-615

c 20	11	91.7	341	10	US-09-783-590-6885	Sequence 6885, Ap
c 21	11	91.7	350	10	US-09-815-343-1513	Sequence 1513, Ap
c 22	11	91.7	385	10	US-09-864-761-16659	Sequence 16659, A
c 23	11	91.7	405	9	US-09-738-626-2571	Sequence 2571, Ap
c 24	11	91.7	411	10	US-09-974-300-7207	Sequence 7207, Ap
c 25	11	91.7	415	10	US-09-834-975-417	Sequence 417, App
c 26	11	91.7	416	12	US-10-044-090-425	Sequence 425, App
c 27	11	91.7	434	10	US-09-960-352-1726	Sequence 1726, Ap
c 28	11	91.7	462	10	US-09-867-701-2628	Sequence 2628, Ap
c 29	11	91.7	474	9	US-10-025-380-1073	Sequence 1073, Ap
c 30	11	91.7	474	10	US-09-922-217-1073	Sequence 1073, Ap
c 31	11	91.7	474	10	US-09-833-263-1073	Sequence 1073, Ap
c 32	11	91.7	475	10	US-09-864-761-4882	Sequence 4882, Ap
c 33	11	91.7	481	10	US-09-998-598-25	Sequence 25, Appl
c 34	11	91.7	503	10	US-09-920-300A-1057	Sequence 1057, Ap
c 35	11	91.7	503	12	US-10-033-528-1057	Sequence 1057, Ap
c 36	11	91.7	551	10	US-09-962-832-3	Sequence 3, Appl
c 37	11	91.7	555	10	US-09-815-343-1539	Sequence 1539, Ap
c 38	11	91.7	555	10	US-09-998-598-1218	Sequence 1218, Ap
c 39	11	91.7	588	10	US-09-815-343-280	Sequence 280, App
c 40	11	91.7	609	10	US-09-815-242-6509	Sequence 6509, Ap
c 41	11	91.7	614	9	US-10-025-380-1075	Sequence 1075, Ap
c 42	11	91.7	614	10	US-09-922-217-1075	Sequence 1075, Ap
c 43	11	91.7	614	10	US-09-833-263-1075	Sequence 1075, Ap
c 44	11	91.7	666	9	US-09-738-626-706	Sequence 706, App
c 45	11	91.7	690	9	US-09-821-877-6	Sequence 6, Appl

ALIGNMENTS

RESULT 1
US-09-560-863-114
; Sequence 114, Application US/09560863
; Patent No. US20020110809A1
; GENERAL INFORMATION:
; APPLICANT: Nehls, Michael C.
; APPLICANT: Zambrowicz, Brian
; APPLICANT: Sands, Arthur T.
; TITLE OF INVENTION: NO. US20020110809A1el Human Polynucleotides and the
; FILE REFERENCE: LEX-0018-USA
; CURRENT APPLICATION NUMBER: US/09/560,863
; PRIOR FILING DATE: 1999-04-28
; PRIOR APPLICATION NUMBER: US 60/132,408
; NUMBER OF SEQ ID NOS: 1008
; SOFTWARE: FASTSEQ for Windows Version 4.0
; SEQ ID NO 114
; LENGTH: 355
; TYPE: DNA
; ORGANISM: homo sapiens
US-09-560-863-114

Query Match 100.0%; Score 12; DB 10; Length 355;
Best Local Similarity 100.0%; Pred. No. 2.1e+02;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GTCGTGACGCA 12
Db 40 GTCGTGACGCA 51

RESULT 2

US-09-917-800A-992
; Sequence 992, Application US/09917800A
; Patent No. US20020119462A1
; GENERAL INFORMATION:
; APPLICANT: Mendrick, Donna
; APPLICANT: Porter, Mark
; APPLICANT: Johnson, Kory
; APPLICANT: Castle, Arthur
; APPLICANT: Elashoff, Michael

APPLICANT: Gene Logic, Inc.
; TITLE OF INVENTION: Molecular Toxicology Modeling
; FILE REFERENCE: 44921-5038-US
; CURRENT APPLICATION NUMBER: US/09/917,800A
; CURRENT FILING DATE: 2001-07-31
; PRIOR APPLICATION NUMBER: US 60/222,040
; PRIOR FILING DATE: 2000-07-31
; PRIOR APPLICATION NUMBER: US 60/222,880
; PRIOR FILING DATE: 2000-11-02
; PRIOR APPLICATION NUMBER: US 60/290,029
; PRIOR FILING DATE: 2001-05-11
; PRIOR APPLICATION NUMBER: US 60/290,645
; PRIOR FILING DATE: 2001-05-15
; PRIOR APPLICATION NUMBER: US 60/292,336
; PRIOR FILING DATE: 2001-05-22
; PRIOR APPLICATION NUMBER: US 60/295,798
; PRIOR FILING DATE: 2001-06-06
; PRIOR APPLICATION NUMBER: US 60/297,457
; PRIOR FILING DATE: 2001-06-13
; PRIOR APPLICATION NUMBER: US 60/298,884
; PRIOR FILING DATE: 2001-06-19
; PRIOR APPLICATION NUMBER: US 60/303,459
; PRIOR FILING DATE: 2001-07-09
; NUMBER OF SEQ ID NOS: 1740
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 992
; LENGTH: 575
; TYPE: DNA
; ORGANISM: Rattus norvegicus
; FEATURE:
; OTHER INFORMATION: Genbank Accession No. US20020119462A1 A1176942
US-09-917-800A-992

Query Match 100.0%; Score 12; DB 10; Length 575;
Best Local Similarity 100.0%; Pred. No. 2.le+02;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GTCGTGCAGGCA 12
|||||

Db 68 GTCGTGCAGGCA 79

RESULT 3
US-09-731-872-80
; Sequence 80, Application US/09731872
; Patent No. US20020102604A1
; GENERAL INFORMATION:
; APPLICANT: Dumas Milne Edwards, Jean Baptiste
; APPLICANT: Bougueleret, Lydie
; APPLICANT: Jobert, Severin
; TITLE OF INVENTION: FULL-LENGTH HUMAN cDNAs ENCODING POTENTIALLY SECRETED PROTEINS
; FILE REFERENCE: 78.US3.REG
; CURRENT APPLICATION NUMBER: US/09/731,872
; CURRENT FILING DATE: 2000-12-07
; PRIOR APPLICATION NUMBER: US 60/169,629
; PRIOR FILING DATE: 1999-12-08
; PRIOR APPLICATION NUMBER: US 60/187,470
; PRIOR FILING DATE: 2000-03-06
; NUMBER OF SEQ ID NOS: 482
; SOFTWARE: Patent.pm
; SEQ ID NO 80
; LENGTH: 1351
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 259..831
; NAME/KEY: sig peptide
; LOCATION: 259..375
; OTHER INFORMATION: Von Heijne matrix
; OTHER INFORMATION: score 5.809301698725
; OTHER INFORMATION: seq FCVCVIAIGVVQA/LI
US-09-731-872-80

Query Match 100.0%; Score 12; DB 10; Length 1351;
Best Local Similarity 100.0%; Pred. No. 2.le+02;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GTCGTGCAGGCA 12
|||||

Db 364 GTCGTGCAGGCA 375

RESULT 4
US-09-731-872-139
; Sequence 139, Application US/09731872
; Patent No. US20020102604A1
; GENERAL INFORMATION:
; APPLICANT: Dumas Milne Edwards, Jean Baptiste
; APPLICANT: Bougueleret, Lydie
; APPLICANT: Jobert, Severin
; TITLE OF INVENTION: FULL-LENGTH HUMAN cDNAs ENCODING POTENTIALLY SECRETED PROTEINS
; FILE REFERENCE: 78.US3.REG
; CURRENT APPLICATION NUMBER: US/09/731,872
; CURRENT FILING DATE: 2000-12-07
; PRIOR APPLICATION NUMBER: US 60/169,629
; PRIOR FILING DATE: 1999-12-08
; PRIOR APPLICATION NUMBER: US 60/187,470
; PRIOR FILING DATE: 2000-03-06
; NUMBER OF SEQ ID NOS: 482
; SOFTWARE: Patent.pm
; SEQ ID NO 139
; LENGTH: 1474
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 16..471
; NAME/KEY: sig peptide
; LOCATION: 16..93
; OTHER INFORMATION: Von Heijne matrix
; OTHER INFORMATION: score 5.809301698725
; OTHER INFORMATION: seq FCVCVIAIGVVQA/LI
US-09-731-872-139

Query Match 100.0%; Score 12; DB 10; Length 1474;
Best Local Similarity 100.0%; Pred. No. 2e+02;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GTCGTGCAGGCA 12
|||||

Db 82 GTCGTGCAGGCA 93

RESULT 5
US-09-915-582-16
; Sequence 16, Application US/09915582
; Patent No. US20020120103A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: 17 Human Secreted Proteins
; FILE REFERENCE: PS723PI
; CURRENT APPLICATION NUMBER: US/09/915,582
; CURRENT FILING DATE: 2001-07-27
; PRIOR APPLICATION NUMBER: PCT/US01/01431
; PRIOR FILING DATE: 2001-01-17
; PRIOR APPLICATION NUMBER: 60/179,065
; PRIOR FILING DATE: 2000-01-31
; PRIOR APPLICATION NUMBER: 60/180,628
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: 60/231,968
; PRIOR FILING DATE: 2000-09-12
; NUMBER OF SEQ ID NOS: 97
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 16
; LENGTH: 1677

; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-915-582-16

Query Match 100.0%; Score 12; DB 10; Length 1677;
Best Local Similarity 100.0%; Pred. No. 2e+02;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GTCGTGCAGGCA 12
|||||
Db 200 GTCGTGCAGGCA 211

RESULT 6

US-09-799-777-104
; Sequence 104, Application US/09799777
; Patent No. US20020091244A1
; GENERAL INFORMATION:

APPLICANT: Lal, Preeti
Hillman, Jennifer L.
Corley, Neil C.
Guegler, Karl J.
Baugh, Mariah
Sather, Susan
Shah, Purvi
TITLE OF INVENTION: HUMAN SIGNAL PEPTIDE-CONTAINING PROTEINS
NUMBER OF SEQUENCES: 154
CORRESPONDENCE ADDRESS:
ADDRESSEE: INCYTE PHARMACEUTICALS, INC.
STREET: 3174 PORTER DRIVE
CITY: PALO ALTO
STATE: CALIFORNIA
COUNTRY: USA
ZIP: 94304

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Word Perfect 6.1 for Windows/MS-DOS 6.2
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/799,777
FILING DATE: 06-Mar-2001
CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US/09/002,485

FILING DATE: <Unknown>

ATTORNEY/AGENT INFORMATION:

NAME: BILLINGS, LUCY J.

REGISTRATION NUMBER: 36,749

REFERENCE/DOCKET NUMBER: PF-0459 US

TELECOMMUNICATION INFORMATION:

TELEPHONE: (650) 855-0555

TELEFAX: (650) 845-4166

INFORMATION FOR SEQ ID NO: 104:

SEQUENCE CHARACTERISTICS:

LENGTH: 1802 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

IMMEDIATE SOURCE:

LIBRARY: PROSTUT08

CLONE: 1653112

SEQUENCE DESCRIPTION: SEQ ID NO: 104 :

US-09-799-777-104

Query Match 100.0%; Score 12; DB 10; Length 1802;
Best Local Similarity 100.0%; Pred. No. 2e+02;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GTCGTGCAGGCA 12
|||||
Db 583 GTCGTGCAGGCA 594

RESULT 7

US-09-738-626-3487
; Sequence 3487, Application US/09738626
; Publication No. US20020197605A1
; GENERAL INFORMATION:

APPLICANT: NAKAGAWA, SATOSHI

APPLICANT: MIZOGUCHI, HIROSHI

APPLICANT: ANDO, SEIKO

APPLICANT: HAYASHI, MIKIRO

APPLICANT: OCHIALI, KEIKO

APPLICANT: YOKOI, HARUHIKO

APPLICANT: TATEISHI, NAKKO

APPLICANT: SENOH, AKIHIRO

APPLICANT: IKEDA, MASATO

APPLICANT: OZAKI, AKIO

TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES

FILE REFERENCE: 249-125

CURRENT APPLICATION NUMBER: US/09/738,626

CURRENT FILING DATE: 2000-12-18

PRIOR APPLICATION NUMBER: JP 99/377484

PRIOR FILING DATE: 1999-12-16

PRIOR APPLICATION NUMBER: JP 00/159162

PRIOR FILING DATE: 2000-04-07

PRIOR APPLICATION NUMBER: JP 00/280988

PRIOR FILING DATE: 2000-08-03

NUMBER OF SEQ ID NOS: 7059

SOFTWARE: PatentIn ver. 3.0

SEQ ID NO 3487

LENGTH: 1845

TYPE: DNA

ORGANISM: Corynebacterium glutamicum

US-09-738-626-3487

Query Match

Best Local Similarity 100.0%; Score 12; DB 9; Length 1845;

Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GTCGTGCAGGCA 12

Db 1139 GTCGTGCAGGCA 1150

RESULT 8

US-09-915-582-34

; Sequence 34, Application US/09915582

; Patent No. US20020120103A1

; GENERAL INFORMATION:

APPLICANT: Rosen et al.

TITLE OF INVENTION: 17 Human Secreted Proteins

FILE REFERENCE: PS723p1

CURRENT APPLICATION NUMBER: US/09/915,582

CURRENT FILING DATE: 2001-07-27

PRIOR APPLICATION NUMBER: PCT/US01/01431

PRIOR FILING DATE: 2001-01-17

PRIOR APPLICATION NUMBER: 60/179,065

PRIOR FILING DATE: 2000-01-31

PRIOR APPLICATION NUMBER: 60/180,628

PRIOR FILING DATE: 2000-02-04

PRIOR APPLICATION NUMBER: 60/231,968

PRIOR FILING DATE: 2000-09-12

NUMBER OF SEQ ID NOS: 97

SOFTWARE: PatentIn Ver. 2.0

SEQ ID NO 34

LENGTH: 1848

TYPE: DNA

ORGANISM: Homo sapiens

FEATURE:

NAME/KEY: SITE

LOCATION: (1796)

OTHER INFORMATION: n equals a,t,g, or c

US-09-915-582-34

Query Match 100.0%; Score 12; DB 10; Length 1848;
Best Local Similarity 100.0%; Pred. No. 2e+02;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GTCGTGCAGGCA 12
|||||
Db 378 GTCGTGCAGGCA 389

RESULT 9

US-09-845-583-1/c
; Sequence 1, Application US/09845583
; Patent No. US20020142954A1
; GENERAL INFORMATION:
; APPLICANT: Burgeson, Robert
; APPLICANT: Brunken, William Joseph
; APPLICANT: Champilaud, Marie-France
; APPLICANT: Hunter, Dale
; TITLE OF INVENTION: LAMININ 15 AND USES THEREOF
; FILE REFERENCE: 10287-056001
; CURRENT APPLICATION NUMBER: US/09/845,583
; CURRENT FILING DATE: 2001-04-30
; PRIOR APPLICATION NUMBER: US 60/200,863
; PRIOR FILING DATE: 2000-05-01
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 11009
; TYPE: DNA
; ORGANISM: Mus musculus
US-09-845-583-1

Query Match 100.0%; Score 12; DB 10; Length 11009;
Best Local Similarity 100.0%; Pred. No. 2e+02;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GTCGTGCAGGCA 12
|||||
Db 9731 GTCGTGCAGGCA 9720

RESULT 10

US-09-867-701-6217/c
; Sequence 6217, Application US/09867701
; Patent No. US20020132237A1
; GENERAL INFORMATION:
; APPLICANT: Aglate, Paul A.
; APPLICANT: Jones, Robert
; APPLICANT: Harlocker, Susan L.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; FILE REFERENCE: 210121.497
; CURRENT APPLICATION NUMBER: US/09/867,701
; CURRENT FILING DATE: 2001-05-29
; NUMBER OF SEQ ID NOS: 10912
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 6217
; LENGTH: 118
; TYPE: DNA
; ORGANISM: Homo sapien
US-09-867-701-6217

Query Match 91.7%; Score 11; DB 10; Length 118;
Best Local Similarity 100.0%; Pred. No. 8.3e+02;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GTCGTGCAGGC 11
|||||
Db 75 GTCGTGCAGGC 65

RESULT 11

US-09-815-343-1511/c

; Sequence 1511, Application US/09815343
; Patent No. US20010055596A1
; GENERAL INFORMATION:
; APPLICANT: Meagher, Madeleine
; APPLICANT: Xu, Jiangchun E.
; APPLICANT: King, Gordon E.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THERAPY AND
; FILE REFERENCE: 210121.504
; CURRENT APPLICATION NUMBER: US/09/815,343
; CURRENT FILING DATE: 2001-03-22
; NUMBER OF SEQ ID NOS: 1556
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1511
; LENGTH: 126
; TYPE: DNA
; ORGANISM: Homo sapien
; FEATURE:
; NAME/KEY: misc.feature
; LOCATION: (1)...(126)
; OTHER INFORMATION: n = A,T,C or G
US-09-815-343-1511

Query Match 91.7%; Score 11; DB 10; Length 126;
Best Local Similarity 100.0%; Pred. No. 8.3e+02;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GTCGTGCAGGC 11
|||||
Db 76 GTCGTGCAGGC 66

RESULT 12

US-09-900-714A-3
; Sequence 3, Application US/09900714A
; Patent No. US20020162133A1
; GENERAL INFORMATION:
; APPLICANT: Allen, Keith D.
; TITLE OF INVENTION: TRANSGENIC MICE CONTAINING MGLUR8
; FILE REFERENCE: R-657
; CURRENT APPLICATION NUMBER: US/09/900,714A
; CURRENT FILING DATE: 2001-07-06
; PRIOR APPLICATION NUMBER: US 60/216,252
; PRIOR FILING DATE: 2000-07-06
; PRIOR APPLICATION NUMBER: US 60/221,490
; PRIOR FILING DATE: 2000-07-27
; PRIOR APPLICATION NUMBER: US 60/262,138
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/300,928
; PRIOR FILING DATE: 2000-07-26
; NUMBER OF SEQ ID NOS: 3
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 200
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Targeting vector
US-09-900-714A-3

Query Match 91.7%; Score 11; DB 9; Length 200;
Best Local Similarity 100.0%; Pred. No. 8.3e+02;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 TCGTGCAGGCA 12
|||||
Db 61 TCGTGCAGGCA 71

RESULT 13

US-09-923-876-557
; Sequence 557, Application US/09923876

```

; Patent No. US20020013958A1
; GENERAL INFORMATION:
; APPLICANT: Lalgudi, Raghunath V.
; APPLICANT: Kamigaki, Laura Y. (Ito)
; APPLICANT: Sherman, Bradley K.
; TITLE OF INVENTION: POLYNUCLEOTIDES AND POLYPEPTIDES DERIVED FROM CORN SEEDLING
; FILE REFERENCE: PL-0012-1 CON
; CURRENT APPLICATION NUMBER: US/09/923,876
; CURRENT FILING DATE: 2001-08-06
; PRIOR APPLICATION NUMBER: 09/298,329
; PRIOR FILING DATE: 1999-04-21
; PRIOR APPLICATION NUMBER: 60/085,331
; PRIOR FILING DATE: 1998-05-05
; NUMBER OF SEQ ID NOS: 6332
; SOFTWARE: PERL Program
; SEQ ID NO 557
; LENGTH: 247
; TYPE: DNA
; ORGANISM: Zea mays
; FEATURE:
; NAME/KEY: misc-feature
; OTHER INFORMATION: Incyte ID No. US20020013958A1 700157144H1
US-09-923-876-557

```

```

Query Match          91.7%; Score 11; DB 10; Length 247;
Best Local Similarity 100.0%; Pred. No. 8.3e+02;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 2 TCGTGCAGGCA 12
    |||||
Db 170 TCGTGCAGGCA 180

```

RESULT 14

```

US-09-960-352-10276/c
; Sequence 10276, Application US/09960352
; Patent No. US20020137139A1
; GENERAL INFORMATION:
; APPLICANT: Warren, Wesley C.
; APPLICANT: Tao, Nengbing
; APPLICANT: Byatt, John C.
; APPLICANT: Mathialagan, Nagappan
; TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND
; FILE REFERENCE: 16511.006/37-21(10298)C
; CURRENT APPLICATION NUMBER: US/09/960,352
; CURRENT FILING DATE: 2001-09-24
; NUMBER OF SEQ ID NOS: 15112
; SEQ ID NO 10276
; LENGTH: 274
; TYPE: DNA
; ORGANISM: Bos taurus
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (19)
; OTHER INFORMATION: unsure at all n locations
; OTHER INFORMATION: Clone ID: 44-LIB3058-056-Q1-K1-C8
US-09-960-352-10276

```

```

Query Match          91.7%; Score 11; DB 10; Length 274;
Best Local Similarity 100.0%; Pred. No. 8.2e+02;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 1 GTCGTGCAGGC 11
    |||||
Db 57 GTCGTGCAGGC 47

```

RESULT 15

```

US-09-841-157A-2
; Sequence 2, Application US/09841157A
; Publication No. US20020192648A1
; GENERAL INFORMATION:

```

```

; APPLICANT: NISHIGAKI, KOICHI
; APPLICANT: TAKASAWA, TSUTOMU
; APPLICANT: HAMANO, KEIICHI
; TITLE OF INVENTION: METHODS OF IDENTIFYING AN ORGANISM BASED ON ITS GENOTYPE
; FILE REFERENCE: 12637/P66602USO
; CURRENT APPLICATION NUMBER: US/09/841,157A
; CURRENT FILING DATE: 2001-04-25
; NUMBER OF SEQ ID NOS: 44
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 286
; TYPE: DNA
; ORGANISM: Unknown Organism
; FEATURE:
; OTHER INFORMATION: Description of Unknown Organism: Illustrative standard
; OTHER INFORMATION: DNA
US-09-841-157A-2

```

```

Query Match          91.7%; Score 11; DB 9; Length 286;
Best Local Similarity 100.0%; Pred. No. 8.2e+02;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 2 TCGTGCAGGCA 12
    |||||
Db 180 TCGTGCAGGCA 190

```

```

Search completed: January 4, 2003, 01:06:18
Job time : 6.05047 secs

```


GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 3, 2003, 22:52:46 ; Search time 216.151 Seconds
(without alignments)
3231.380 Million cell updates/sec

Title: US-09-787-562-11

Perfect score: 24

Sequence: 1 tctagtgtcgtgcaggcatcgtagt 24

Scoring table: IDENTITY_NUC

Gapop 10.0, Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl.*

1: gb_ba.*

2: gb_htg.*

3: gb_in.*

4: gb_om.*

5: gb_ov.*

6: gb_pat.*

7: gb_ph.*

8: gb_pl.*

9: gb_pr.*

10: gb_ro.*

11: gb_sts.*

12: gb_sy.*

13: gb_un.*

14: gb_vi.*

15: em_ba.*

16: em_fun.*

17: em_hum.*

18: em_in.*

19: em_mu.*

20: em_om.*

21: em_or.*

22: em_ov.*

23: em_pat.*

24: em_ph.*

25: em_pl.*

26: em_ro.*

27: em_sts.*

28: em_un.*

29: em_vi.*

30: em_htg_hum.*

31: em_htg_inv.*

32: em_htg_other.*

33: em_htg_mus.*

34: em_htg_pin.*

35: em_htg_rod.*

36: em_htg_mam.*

37: em_htg_vrt.*

38: em_sy.*

39: em_higo_hum.*

40: em_higo_mus.*

41: em_higo_other.*

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	24	100.0	24	6	AX023669	Sequence
2	24	100.0	237	6	AX023667	Sequence
3	18.8	78.3	183342	2	AC129440	Rattus no
4	18.2	75.8	64672	2	AC126356	Homo sapi
5	18.2	75.8	103432	2	AC094981	Rattus no
6	18.2	75.8	182943	9	AC016251	Homo sapi
7	18.2	75.8	200078	10	AL591064	Mouse DNA
8	18.2	75.8	229662	2	AC125109	Mus muscu
9	17.8	74.2	1768	1	HSAMP03S24	Human AMP d
10	17.8	74.2	40520	1	AC090969	Staphyloc
11	17.8	74.2	69320	2	AC113035	Mus muscu
12	17.8	74.2	152631	2	AC048373	Homo sapi
13	17.8	74.2	156507	9	AC021914	Homo sapi
14	17.8	74.2	177429	9	AP002378	Homo sapi
15	17.8	74.2	177488	9	AP000760	Homo sapi
16	17.8	74.2	204140	9	AC010976	Homo sapi
17	17.8	74.2	272850	1	AP004828	Staphyloc
18	17.8	74.2	291150	1	AP003135	Staphyloc
19	17.8	74.2	342600	1	AP003363	Human mRNA
20	17.6	73.3	6237	9	D86978	Human mRNA
21	17.6	73.3	82642	2	AC111397	Rattus no
22	17.6	73.3	110000	2	AC095071_3	Continuation (4 of
23	17.6	73.3	168089	2	AC113849	Rattus no
24	17.6	73.3	183977	2	AC097956	Rattus no
25	17.6	73.3	197837	9	AC093107	Homo sapi
26	17.6	73.3	207585	2	AC073779	Mus muscu
27	17.6	73.3	235218	2	AC074208	Mus muscu
28	17.6	73.3	247593	2	AC103319	Rattus no
29	17.4	72.5	72	6	AX023690	Sequence
30	17.4	72.5	72	6	AX023692	Sequence
31	17.2	71.7	1676	1	ECONEUC	M84026 E.coli prot
32	17.2	71.7	59634	9	AC062028	Homo sapi
33	17.2	71.7	61796	2	AC020875	Mus muscu
34	17.2	71.7	95560	2	AC127620	Rattus no
35	17.2	71.7	102345	9	AL445230	Human DNA
36	17.2	71.7	120739	2	AL807782	Homo sapi
37	17.2	71.7	170677	2	AL138932	Homo sapi
38	17.2	71.7	170965	2	AC108288	Rattus no
39	17.2	71.7	175758	9	AC098825	Homo sapi
40	17.2	71.7	175957	2	AL691501	Mus muscu
41	17.2	71.7	180418	9	AC093875	Homo sapi
42	17.2	71.7	183338	2	AC120970	Rattus no
43	17.2	71.7	185425	2	AC124837	Rattus no
44	17.2	71.7	185532	10	AC121883	Mus muscu
45	17.2	71.7	196203	9	AL355355	Human DNA

ALIGNMENTS

RESULT 1
AX023669

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

AX023669 Sequence 11 from Patent WO0017371. 24 bp DNA Linear PAT 16-SEP-2000

AX023669

AX023669.1 GI:10184030

synthetic construct.

synthetic construct

artificial sequences.

1 (bases 1 to 24)

Binley,K.M. and Naylor,S.

Polynucleotide constructs and uses thereof

Patent: WO 0017371-A 11 30-MAR-2000;

BINLEY KATIE MARY (GB) ; NAYLOR STUART (GB) ; OXFORD BIOMEDICA LTD

Pred. No. is the number of results predicted by chance to have a

```

(GB)
FEATURES
  source
    Location/Qualifiers
      1..24
        /organism="synthetic construct"
        /db_xref="taxon:32630"
        /note="Spacer"
BASE COUNT
  4 a 5 c 7 g 8 t
ORIGIN
  Query Match 100.0%; Score 247. DB 6; Length 24;
  Best Local Similarity 100.0%; Pred. No. 0.045;
  Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 TCTAGTGTGTCGAGGCATCTAGT 24
  |||||
Db 1 TCTAGTGTGTCGAGGCATCTAGT 24
  |||||

RESULT 2
AX023667
LOCUS AX023667 237 bp DNA linear PAT 15-SEP-2000
DEFINITION Sequence 9 from Patent WO0017371.
ACCESSION AX023667
VERSION AX023667.1 GI:10184028
KEYWORDS
  synthetic construct.
  artificial construct.
  synthetic sequences.
  1 (bases 1 to 237)
REFERENCE
  Binley,K.M. and Naylor,S.
  Polynucleotide constructs and uses thereof
  Patent: WO 0017371-A 9 30-MAR-2000;
  TITLE WO 0017371-A 9 30-MAR-2000;
  JOURNAL PATENT: WO 0017371-A 9 30-MAR-2000;
  BINGLEY KATIE MARY (GB) ; NAYLOR STUART (GB) ; OXFORD BIOMEDICA LTD
  (GB)
FEATURES
  source
    Location/Qualifiers
      1..237
        /organism="synthetic construct"
        /db_xref="taxon:32630"
        /note="OBHrel1"
BASE COUNT
  43 a 82 c 56 g 56 t
ORIGIN
  Query Match 100.0%; Score 24; DB 6; Length 237;
  Best Local Similarity 100.0%; Pred. No. 0.041;
  Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 TCTAGTGTGTCGAGGCATCTAGT 24
  |||||
Db 25 TCTAGTGTGTCGAGGCATCTAGT 48
  |||||

RESULT 3
AC129440
LOCUS AC129440 183342 bp DNA linear HTG 30-JUL-2002
DEFINITION Rattus norvegicus clone CH230-112K14, *** SEQUENCING IN PROGRESS
  ***. 65 unordered pieces.
ACCESSION AC129440
VERSION AC129440.1 GI:22004125
KEYWORDS HTG; HTGS_PHASE1.
SOURCE Norway rat.
  Rattus norvegicus
  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
  Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
  Rattus.
  1 (bases 1 to 183342)
REFERENCE
  Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
  Alsbrooks,S.L., Anaratunge,H.C., Are,J.R., Ayele,M., Banks,T.,
  Barbara,J., Benton,J., Binage,K., Blankenburg,K., Bonnin,D.,
  Bouck,J., Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P.,
  Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C.,
  Caron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,
  Chen,G., Chen,K., Chen,Z., Chowdhry,I., Christopoulos,C.,
  Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.K., David,R.,
  Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A.,
  Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H.,
  Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J.,
  Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M.,
  Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P.,
  Galis,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R.,
  Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K.,
  Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., Hernandez,J.,
  Hernandez,O., Hodgson,A., Hogues,M., Holloway,C., Hollins,B.,
  Homi,F., Howard,S., Huber,J., Hulyk,S., Hume,J., Jackson,L.E.,
  Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S.,
  Karlsson,E., Kelly,S., Khan,U., King,L., Korvan,J., Kovar,C.,
  Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L.,
  Li,J., Li.Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W., Loulseged,H.,
  Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J.,
  Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E.,
  Massey,B., Mawhney,E., McLeod,M.P., Meador,M., Mei,G., Mettler,M.,
  Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S.,
  Moser,M., Neal,D., Newton,J., Newton,N., Nguyen,A., Nguyen,S.,
  Nguyen,N., Nickerson,E., Nwokenwo,S., Ogih,M., Okwuonu,G.,
  Oragunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L.,
  Peters,L., Picken,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y.,
  Rives,M., Rojas,A., Rojibokan,I., Rolfe,M., Ruiz,S., Savery,G.,
  Scherer,S., Scott,G., Shen,H., Shoostari,N., Sisson,I.,
  Sodergren,E., Sonaik,T., Sparks,A., Stanley,H., Stone,H.,
  Sutton,A., Svatek,A., Tabot,P., Tamerisa,A., Tamerisa,K., Tang,H.,
  Tansey,J., Taylor,C., Taylor,T., Telford,B., Thomas,N., Thomas,S.,
  Usmani,K., Vasquez,L., Vera,V., Villalón,D., Vinson,R., Wang,Q.,
  Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S.,
  Williams,G., Williamson,A., Wleczyk,R., Wooden,S., Worley,K.,
  Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,
  Weinstein,G., and Gibbs,R.
  Direct Submission
  Unpublished
  2 (bases 1 to 183342)
  Worley,K.C.
  Direct Submission
  Submitted (30-JUL-2002) Human Genome Sequencing Center, Department
  of Molecular and Human Genetics, Baylor College of Medicine, One
  Baylor Plaza, Houston, TX 77030, USA
  ----- Genome Center
  Center: Baylor College of Medicine
  Center code: BCM
  Web site: http://www.hgsc.bcm.tmc.edu/
  Contact: hgsc-help@bcm.tmc.edu
  ----- Project Information
  Center project name: GKTY
  Center clone name: CH230-112K14
  ----- Summary Statistics
  Sequencing vector: Plasmid;
  Chemistry: Dye-terminator Big Dye; 100% of reads
  Assembly program: Phrap; version 0.990329
  Consensus quality: 130157 bases at least Q40
  Consensus quality: 135959 bases at least Q30
  Consensus quality: 140433 bases at least Q20
  -----
  * NOTE: Estimated insert size may differ from sequence length
  (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).
  * NOTE: This is a 'working draft' sequence. It currently
  * consists of 65 contigs. The true order of the pieces
  * is not known and their order in this sequence record is
  * arbitrary. Gaps between the contigs are represented as
  * runs of N, but the exact sizes of the gaps are unknown.
  * This record will be updated with the finished sequence
  * as soon as it is available and the accession number will
  * be preserved.
  * 1 1012: contig of 1012 bp in length
  * 1013 1112: gap of unknown length
  * 1113 2240: contig of 1028 bp in length
  * 2241 3744: gap of unknown length
  * 3745 4961: contig of 1504 bp in length
  * 3844: gap of unknown length
  * 4961: contig of 1117 bp in length

```

Mon Jan 6 15:20:20 2003

```

* 4962 5061: gap of unknown length
* 5062 6356: contig of 1295 bp in length
* 6456: gap of unknown length
* 6457 8020: contig of 1564 bp in length
* 8021 8120: gap of unknown length
* 8121 9729: contig of 1609 bp in length
* 9730 9829: gap of unknown length
* 9830 10968: contig of 1139 bp in length
* 10969 11068: gap of unknown length
* 11070 12102: contig of 1034 bp in length
* 12103 12202: gap of unknown length
* 12204 13464: contig of 1262 bp in length
* 13465 13564: gap of unknown length
* 13565 15019: contig of 1455 bp in length
* 15020 15119: gap of unknown length
* 15120 16277: contig of 1158 bp in length
* 16278 16377: gap of unknown length
* 16379 17598: contig of 1221 bp in length
* 17599 17698: gap of unknown length
* 17700 19126: contig of 1428 bp in length
* 19127 19226: gap of unknown length
* 19228 20838: contig of 1612 bp in length
* 20839 22282: contig of 1344 bp in length
* 22283 23806: contig of 1424 bp in length
* 23807 23906: gap of unknown length
* 23907 25005: contig of 1099 bp in length
* 25006 25105: gap of unknown length
* 25106 26897: contig of 1792 bp in length
* 26898 26997: gap of unknown length
* 26998 29011: contig of 2014 bp in length
* 29012 29111: gap of unknown length
* 29112 30662: contig of 1551 bp in length
* 30663 30762: gap of unknown length
* 30763 32775: contig of 2013 bp in length
* 32776 32875: gap of unknown length
* 32877 34108: contig of 1233 bp in length
* 34109 34208: gap of unknown length
* 34209 35893: contig of 1685 bp in length
* 35894 35993: gap of unknown length
* 35994 38034: contig of 2041 bp in length
* 38035 38134: gap of unknown length
* 38135 39356: contig of 1222 bp in length
* 39357 39456: gap of unknown length
* 39457 41734: contig of 2278 bp in length
* 41735 41834: gap of unknown length
* 41836 43106: contig of 1272 bp in length
* 43107 43206: gap of unknown length
* 43207 44719: contig of 1513 bp in length
* 44720 46402: contig of 1583 bp in length
* 46403 46502: gap of unknown length
* 46503 48677: contig of 2175 bp in length
* 48678 48777: gap of unknown length
* 48779 50549: contig of 1772 bp in length
* 50550 50649: gap of unknown length
* 50650 51975: contig of 1326 bp in length
* 51976 52075: gap of unknown length
* 52076 54622: contig of 2547 bp in length
* 54623 54722: gap of unknown length
* 54723 57428: contig of 2706 bp in length
* 57429 57528: gap of unknown length
* 57529 60703: contig of 3175 bp in length
* 60704 60803: gap of unknown length
* 60804 62969: contig of 2166 bp in length
* 62970 63069: gap of unknown length
* 63070 65319: contig of 2250 bp in length
* 65320 65419: gap of unknown length
* 65420 68626: contig of 3207 bp in length
* 68627 71376: gap of unknown length
* 71377 71476: contig of 2650 bp in length
* 71477 73563: contig of 2087 bp in length
* 73564 73663: gap of unknown length
* 73664 76959: contig of 3295 bp in length
* 76960 77059: gap of unknown length
* 77060 79871: contig of 2812 bp in length
* 79872 79971: gap of unknown length
* 79972 83051: contig of 3080 bp in length
* 83052 83151: gap of unknown length
* 83152 86495: contig of 3345 bp in length
* 86496 86595: gap of unknown length
* 86596 89681: contig of 3086 bp in length
* 89682 89781: gap of unknown length
* 89782 92489: contig of 2708 bp in length
* 92490 92589: gap of unknown length
* 92590 95401: contig of 2812 bp in length
* 95402 95501: gap of unknown length
* 95502 98592: contig of 3091 bp in length
* 98593 98693: gap of unknown length
* 98694 101719: contig of 3027 bp in length
* 101720 104542: contig of 2723 bp in length
* 104543 104642: gap of unknown length
* 104643 108398: contig of 3756 bp in length
* 108399 108498: gap of unknown length
* 108499 112294: contig of 3795 bp in length
* 112295 112394: gap of unknown length
* 112395 116431: contig of 4038 bp in length
* 116432 116531: gap of unknown length
* 116532 124038: contig of 7507 bp in length
* 124039 124138: gap of unknown length
* 124139 128957: contig of 4819 bp in length
* 128958 129057: gap of unknown length
* 129058 135305: contig of 6248 bp in length
* 135306 135405: gap of unknown length

Query Match 78.3%; Score 18.8; DB 2; Length 183342;
Best Local Similarity 90.9%; Pred. No. 26;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 TCTAGTGTCTGTCAGGCATCTA 22
||||| | ||||| |||||
Db 120514 TCTAGTACCTGTCAGGCATCTA 120535

RESULT 4
AC126356 Homo sapiens chromosome 17 clone CTD-2541111 map 17, LOW-PASS
LOCUS AC126356 64672 bp DNA linear HTG 30-JUL-2002
DEFINITION AC126356 SEQUENCE SAMPLING.
ACCESSION AC126356.2 GI:22004353
VERSION AC126356.2
KEYWORDS HTG; HTGS_PHASE0.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 64672)
AUTHORS Birren, B., Nussbaum, C. and Lander, E.
TITLE Homo sapiens chromosome 17, clone CTD-2541111
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 64672)
AUTHORS Birren, B., Nussbaum, C., Lander, E., Ali, A., Allen, N., Anderson, S.,
Barna, N., Bastien, V., Bloom, T., Boguslavsky, L., Boukhgalter, B.,
Camarata, J., Chang, J., Chazaro, B., Choepel, Y., Collymore, A.,
Cook, A., Cooke, P., DeArellano, K., Dewar, K., Diaz, J.S., Dodge, S.,
Faro, S., Ferreira, P., FitzGerald, M., Gage, D., Galagan, J.,
Gardyna, S., Gord, S., Graham, L., Grand-Pierre, N., Hagos, B.,
Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C., Kamat, A.,
Karatas, A., Kells, C., Landers, T., Levine, R., Lindblad-Toh, K.,
Liu, G., MacLean, C., Macdonald, P., Major, J., Matthews, C.,
McCarthy, M., Meidirm, J., Meneus, L., Mihova, T., Milenga, V.,
Murphy, T., Naylor, J., Nguyen, C., Nicol, R., Norbu, C., Norman, C.H.,
O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J., Peterson, K.,

```

Phunkhang, P., Pierre, N., Raymond, C., Retta, R., Rise, C., Rogov, P., Roman, J., Roy, A., Schauer, S., Schuback, R., Seaman, S., Severy, P., Smith, C., Spencer, B., Stange-Thomann, N., Stojanovic, N., Talamas, J., Tesfaye, S., Theodore, J., Topham, K., Travers, M., Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.

TITLE
JOURNALREFERENCE
AUTHORS

Submitted (05-JUL-2002) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA

3 (bases 1 to 64672)

Birren, B., Nussbaum, C., Lander, E., Ali, A., Allen, N., Anderson, S., Barna, N., Bastien, V., Bloom, T., Boguslavsky, L., Boukhgalter, B., Canarata, J., Chang, J., Chazaro, P., Choepel, Y., Collymore, A., Cook, A., Cooke, P., Dearellano, K., Dewar, K., Diaz, J., Dodge, S., Faro, S., Ferreira, P., Fitzgerald, M., Gage, D., Galagan, J., Gardyna, S., Gord, S., Graham, L., Grand-Pierre, N., Hagos, B., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C., Kamat, A., Karatas, A., Kells, C., Landers, T., Levine, R., Lindblad-Toh, K., Liu, G., Maclean, C., Macdonald, P., Major, J., Matthews, C., McCarthy, M., Meldrim, J., Meneus, L., Mihova, T., Mlenga, V., Murphy, T., Naylor, J., Nguyen, C., Nicol, R., Norbu, C., Norman, C.H., O'Connor, T., O'Donnell, P., O'Neil, D., Oliver, J., Peterson, K., Phunkhang, P., Pierre, N., Raymond, C., Retta, R., Rise, C., Rogov, P., Roman, J., Roy, A., Schauer, S., Schuback, R., Seaman, S., Severy, P., Smith, C., Spencer, B., Stange-Thomann, N., Stojanovic, N., Talamas, J., Tesfaye, S., Theodore, J., Topham, K., Travers, M., Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.

TITLE
JOURNAL

COMMENT

Submitted (30-JUL-2002) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA

On Jul 30, 2002 this sequence version replaced gi:21699304.

All repeats were identified using RepeatMasker:

Smith, A.F.A. & Green, P. (1996-1997)

<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

----- Genome Center

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WIBR

Web site: <http://www-seq.wi.mit.edu>

Contact: sequence_submissions@genome.wi.mit.edu

----- Project Information

Center project name: L27688

Center clone name: 2541_I_11

* NOTE: This record contains 78 individual
* sequencing reads that have not been assembled into
* contigs. Runs of N are used to separate the reads
* and the order in which they appear is completely
* arbitrary. Low-pass sequence sampling is useful for
* identifying clones that may be gene-rich and allows
* overlap relationships among clones to be deduced.
* However, it should not be assumed that this clone
* will be sequenced to completion. In the event that
* the record is updated, the accession number will
* be preserved.

1 704: contig of 704 bp in length
* 705 804: gap of 100 bp
* 805 1540: contig of 736 bp in length
* 1541 1640: gap of 100 bp
* 1641 2370: contig of 730 bp in length
* 2371 2470: gap of 100 bp
* 2471 3203: contig of 733 bp in length
* 3204 3303: gap of 100 bp
* 3304 4019: contig of 716 bp in length
* 4020 4119: gap of 100 bp
* 4120 4870: contig of 751 bp in length
* 4871 4970: gap of 100 bp
* 4971 5690: contig of 720 bp in length
* 5691 5790: gap of 100 bp
* 5791 6514: contig of 724 bp in length
* 6515 6614: gap of 100 bp
* 6615 7363: contig of 749 bp in length
* 7364 7463: gap of 100 bp

7464 8203: contig of 740 bp in length
* 8204 8303: gap of 100 bp
* 8304 9028: contig of 725 bp in length
* 9029 9128: gap of 100 bp
* 9129 9837: contig of 729 bp in length
* 9838 9957: gap of 100 bp
* 9958 10683: contig of 726 bp in length
* 10684 10783: gap of 100 bp
* 10784 11499: contig of 716 bp in length
* 11500 11599: gap of 100 bp
* 11600 12332: contig of 733 bp in length
* 12333 12432: gap of 100 bp
* 12433 13176: contig of 744 bp in length
* 13177 13276: gap of 100 bp
* 13277 14013: contig of 737 bp in length
* 14014 14113: gap of 100 bp
* 14114 14840: contig of 727 bp in length
* 14841 14940: gap of 100 bp
* 14941 15671: contig of 731 bp in length
* 15672 15771: gap of 100 bp
* 15772 16504: contig of 733 bp in length
* 16505 16604: gap of 100 bp
* 16605 17333: contig of 729 bp in length
* 17334 17433: gap of 100 bp
* 17434 18172: contig of 739 bp in length
* 18173 18272: gap of 100 bp
* 18273 19000: contig of 728 bp in length
* 19001 19100: gap of 100 bp
* 19101 19835: contig of 735 bp in length
* 19836 19935: gap of 100 bp
* 19936 20676: contig of 741 bp in length
* 20677 20776: gap of 100 bp
* 20777 21523: contig of 747 bp in length
* 21524 21623: gap of 100 bp
* 21624 22370: contig of 747 bp in length
* 22371 22470: gap of 100 bp
* 22471 23197: contig of 727 bp in length
* 23198 23297: gap of 100 bp
* 23298 24043: contig of 746 bp in length
* 24044 24143: gap of 100 bp
* 24144 24870: contig of 727 bp in length
* 24871 24970: gap of 100 bp
* 24971 25682: contig of 712 bp in length
* 25683 25782: gap of 100 bp
* 25783 26510: contig of 728 bp in length
* 26511 26610: gap of 100 bp
* 26611 27335: contig of 725 bp in length
* 27336 27435: gap of 100 bp
* 27436 28180: contig of 745 bp in length
* 28181 28280: gap of 100 bp
* 28281 29005: contig of 725 bp in length
* 29006 29105: gap of 100 bp
* 29106 29842: contig of 737 bp in length
* 29843 29942: gap of 100 bp
* 29943 30669: contig of 727 bp in length
* 30670 30769: gap of 100 bp
* 30770 31504: contig of 735 bp in length
* 31505 31604: gap of 100 bp
* 31605 32336: contig of 732 bp in length
* 32337 32436: gap of 100 bp
* 32437 33163: contig of 727 bp in length
* 33164 33263: gap of 100 bp
* 33264 33994: contig of 731 bp in length
* 33995 34094: gap of 100 bp
* 34095 34826: contig of 732 bp in length
* 34827 34926: gap of 100 bp
* 34927 35658: contig of 732 bp in length
* 35659 35758: gap of 100 bp
* 35759 36505: contig of 747 bp in length
* 36506 36605: gap of 100 bp
* 36606 37335: contig of 730 bp in length
* 37336 37435: gap of 100 bp
* 37436 38155: contig of 720 bp in length

10142 11400: contig of 1259 bp in length
 * 11500: gap of unknown length
 * 11501 12525: contig of 1025 bp in length
 * 12625: gap of unknown length
 * 12626 14327: contig of 1702 bp in length
 * 14328 14427: gap of unknown length
 * 14428 15457: contig of 1030 bp in length
 * 15458 15557: gap of unknown length
 * 15558 17456: contig of 1899 bp in length
 * 17457 17556: gap of unknown length
 * 17557 18969: contig of 1413 bp in length
 * 18970 19069: gap of unknown length
 * 19070 20923: contig of 1854 bp in length
 * 20924 21023: gap of unknown length
 * 21024 22611: contig of 1588 bp in length
 * 22612 22711: gap of unknown length
 * 22712 23937: contig of 1226 bp in length
 * 23938 24037: gap of unknown length
 * 24038 25105: contig of 1068 bp in length
 * 25106 25205: gap of unknown length
 * 25206 26779: contig of 1574 bp in length
 * 26780 26879: gap of unknown length
 * 26880 28040: contig of 1161 bp in length
 * 28041 28140: gap of unknown length
 * 28141 29160: contig of 1020 bp in length
 * 29161 29260: gap of unknown length
 * 29261 30709: contig of 1449 bp in length
 * 30710 30809: gap of unknown length
 * 30810 32063: contig of 1254 bp in length
 * 32064 32163: gap of unknown length
 * 32164 33748: contig of 1585 bp in length
 * 33749 33848: gap of unknown length
 * 33849 35545: contig of 1697 bp in length
 * 35546 36645: gap of unknown length
 * 36646 36881: contig of 1236 bp in length
 * 36882 36981: gap of unknown length
 * 36982 38162: contig of 1181 bp in length
 * 38163 38262: gap of unknown length
 * 38263 40247: contig of 1985 bp in length
 * 40248 40347: gap of unknown length
 * 40348 42294: contig of 1947 bp in length
 * 42295 42394: gap of unknown length
 * 42395 44540: contig of 2146 bp in length
 * 44541 44640: gap of unknown length
 * 44641 46066: contig of 1426 bp in length
 * 46067 46166: gap of unknown length
 * 46167 47884: contig of 1718 bp in length
 * 47885 47984: gap of unknown length
 * 47986 49887: contig of 1903 bp in length
 * 49888 49987: gap of unknown length
 * 49989 52297: contig of 2310 bp in length
 * 52298 52397: gap of unknown length
 * 52399 53710: contig of 1313 bp in length
 * 53711 53810: gap of unknown length
 * 53811 55331: contig of 1521 bp in length
 * 55332 55431: gap of unknown length
 * 55432 57410: contig of 1979 bp in length
 * 57411 57510: gap of unknown length
 * 57511 58818: contig of 1308 bp in length
 * 58819 58918: gap of unknown length
 * 58919 60139: contig of 1221 bp in length
 * 60140 60239: gap of unknown length
 * 60240 63089: contig of 2850 bp in length
 * 63090 63189: gap of unknown length
 * 63190 65344: contig of 2155 bp in length
 * 65345 65444: gap of unknown length
 * 65445 67327: contig of 1883 bp in length
 * 67328 67427: gap of unknown length
 * 67428 68486: contig of 1059 bp in length
 * 68487 68586: gap of unknown length
 * 68587 71481: contig of 2895 bp in length
 * 71482 71581: gap of unknown length
 * 71582 74211: contig of 2630 bp in length

74212 74311: gap of unknown length
 * 76312 76314: contig of 2003 bp in length
 * 76315 76414: gap of unknown length
 * 76415 78604: contig of 2190 bp in length
 * 78605 78704: gap of unknown length
 * 78705 80641: contig of 1937 bp in length
 * 80642 80741: gap of unknown length
 * 80742 83308: contig of 2567 bp in length
 * 83309 83408: gap of unknown length
 * 83409 86076: contig of 2668 bp in length
 * 86077 86176: gap of unknown length
 * 86177 89594: contig of 3418 bp in length
 * 89595 89694: gap of unknown length
 * 89695 92732: contig of 3038 bp in length
 * 92733 92832: gap of unknown length
 * 92833 96349: contig of 3517 bp in length
 * 96350 96449: gap of unknown length
 * 96450 100205: contig of 3756 bp in length
 * 100206 103005: gap of unknown length
 * 100306 103432: contig of 3127 bp in length.

Query Match 75.8%; Score 18.2; DB 2; Length 103432;

Best Local Similarity 87.0%; Pred. No. 57; Mismatches 3; Indels 0; Gaps 0;

Matches 20; Conservative 0;

QY 2 CTAGTGTCTGCAGGCATAGT 24

Db 64816 CGAGTGTCTGCAGGCATATAGT 64838

RESULT 6

AC016251

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

AC016251 182943 bp DNA linear PRI 05-FEB-2002
 Homo sapiens chromosome 15, clone RP11-759A24, complete sequence.
 AC016251
 AC016251.9 GI:18139499
 HTG.
 Homo sapiens.
 Homo sapiens.
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 182943)
 Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
 Brown,A., Castle,A., Collinge,M., Collins,S., Collins,A., Cooke,P.,
 DeArelano,K., Dewar,K., Domino,M., Donelan,L., Doyle,M.,
 Ferreira,P., FitzHugh,W., Forrest,C., Funke,R., Gage,D.,
 Galagan,J., Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,
 Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,
 Lechoczy,J., Lieu,C., Locke,K., MacDonald,P., Marquis,N.,
 McEwan,P., McGurk,A., McKernan,K., McLaughlin,J., Melidri,P.,
 Morrow,J., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
 Peterson,K., Pollara,V., Riley,R., Roy,A., Santos,R., Severy,P.,
 Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
 Tesfaye,S., Tirrell,A., Vassiliev,H., Vo,A., Wheeler,J., Wu,X.,
 Wyman,D., Ye,W.J., Zimmer,A. and Zody,M.
 Direct Submission
 Submitted (23-NOV-1999) Whitehead Institute/MIT Center for Genome
 Research, 320 Charles Street, Cambridge, MA 02141, USA
 3 (bases 1 to 182943)

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

AC016251

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

AC016251

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

AC016251

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

AC016251

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

AC016251

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

AC016251

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

AC016251

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

AC016251

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

AC016251

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

AC016251

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

AC016251

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

AC016251

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

AC016251

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

AC016251

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

AC016251

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

TITLE	Source	Location/Qualifiers
JOURNAL	1..182943	/organism="Homo sapiens"
COMMENT	Research, 320 Charles Street, Cambridge, MA 02141, USA	/db_xref="taxon:9606"
	On Jan 12, 2002 this sequence version replaced gi:17646997.	/chromosome="15"
	All repeats were identified using RepeatMasker:	/map="15"
	Smit, A.F.A. & Green, P. (1996-1997)	/clone="RP11-759A24"
	http://ftp.genome.washington.edu/RM/RepeatMasker.html	/clone.lib="RPCI-11 Human Male BAC"
	----- Genome Center	976..998
	Center: Whitehead Institute/ MIT Center for Genome Research	/rpt_family="CCCCCGn"
	Center code: WtBR	/rpt_family="(A)n"
	Web site: http://www.seq.wi.mit.edu	1476..1496
	Contact: sequence.submissions@genome.wi.mit.edu	/rpt_family="CT-rich"
	----- Project Information	2486..2532
	Center project name: L4985	4823..4868
	Center clone name: 759_A_24	/rpt_family="(GAA)n"
	-----	8642..8716
		/rpt_family="GA-rich"
		9933..9970
		/rpt_family="GC-rich"
		10085..10140
		/rpt_family="GC-rich"
		10337..10407
		/rpt_family="G-rich"
		10471..10519
		/rpt_family="CT-rich"
		10549..10615
		/rpt_family="(GGA)n"
		10738..10765
		/rpt_family="GC-rich"
		10849..10892
		/rpt_family="GC-rich"
		10965..11107
		/rpt_family="(CCG)n"
		11665..11700
		/rpt_family="GC-rich"
		11727..11749
		/rpt_family="(CCG)n"
		14989..15045
		/rpt_family="CT-rich"
		19940..19960
		/rpt_family="GC-rich"
		21261..21312
		/rpt_family="CA-rich"
		21586..21616
		/rpt_family="(A)n"
		24233..24266
		/rpt_family="AT-rich"
		25606..25643
		/rpt_family="(TA)n"
		27005..27029
		/rpt_family="(TG)n"
		complement(27423)..27516
		/rpt_family="MIR"
		complement(27627)..27872
		/rpt_family="MIR"
		27965..28115
		/rpt_family="MIR3"
		complement(28429)..28523
		/rpt_family="MIR"
		30255..30295
		/rpt_family="(TG)n"
		31765..31786
		/rpt_family="AT-rich"
		complement(32307)..32555
		/rpt_family="MIR"
		complement(33254)..33370
		/rpt_family="MIR"
		34793..35088
		/rpt_family="AluJb"
		complement(35759)..35827
		/rpt_family="L3"
		complement(35973)..36232
		/rpt_family="L3"
		complement(36271)..36554
		/rpt_family="AluY"
		complement(36695)..36792
		/rpt_family="L3"
		36994..37220
		/rpt_family="MIR"
		37252..37337
		/rpt_family="L2"
		37802..37831
		/rpt_family="AT-rich"
		39233..39358
		/rpt_family="MIR"
		complement(41525)..41648
		/rpt_family="HALib"
		complement(41813)..42129
		/rpt_family="AluJb"
		43185..43345
		/rpt_family="FAM"
		complement(43644)..43767
		/rpt_family="LIM4"
		45981..46111
		/rpt_family="MIR3"
		46392..46470
		/rpt_family="MIR3"
		complement(46490)..46783
		/rpt_family="AluSg1"
		47905..48006
		/rpt_family="L2"
		complement(48460)..48668
		/rpt_family="AluJo"
		complement(48676)..48918
		/rpt_family="AluSq"
		49562..49582
		/rpt_family="AT-rich"
		complement(49874)..49989
		/rpt_family="MIR"
		50272..50550
		/rpt_family="AluSc"
		complement(50799)..51139
		/rpt_family="AluSg"
		51888..51756
		/rpt_family="T-rich"
		complement(51971)..52143


```

/organism="Mus musculus"
/db_xref="taxon:10090"
/chromosome="UNK"
/clone="RP24-232H5"
1..19665
/note="assembly_name:Contig10"
19766..85228
/note="assembly_name:Contig11"
85329..125046
/note="assembly_name:Contig12"
125147..171694
/note="assembly_name:Contig13"
171795..229662
/note="assembly_name:Contig14"
73610 a 42903 c 42981 g 69761 t 407 others
ORIGIN

Query Match 75.8%; Score 18.2; DB 2; Length 229662;
Best Local Similarity 87.0%; Pred. No. 55;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 TCTAGTGTGTCGAGGCATCTAG 23
||||| |||||||
Db 201681 TTTAGTGAAGTCGAGGCATCTAG 201703

RESULT 9
HSAMPD3S24/c
LOCUS HSAMPD3S24 1768 bp DNA linear PRI 11-JUL-1996
DEFINITION Human AMP deaminase (AMPD3) gene, exon 8 and 9.
ACCESSION U29917
VERSION U29917.1 GI:1002652
KEYWORDS 24 of 32
SEGMENT
SOURCE Homo sapiens.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 1768)
AUTHORS Mahake-Zizelman,D.K., Eddy,R., Shows,T.B. and Sabina,R.L.
TITLE Characterization of the human AMPD3 gene reveals that 5' exon
usage is subject to transcriptional control by three tandem
promoters and alternative splicing
Biochim. Biophys. Acta 1306 (1), 75-92 (1996)
96201708
PUBMED 8611627
REFERENCE 2 (bases 1 to 1768)
AUTHORS Sabina,R.L.
TITLE Direct Submission
JOURNAL Submitted (22-JUN-1995) Richard L. Sabina, Biochemistry, Medical
College of Wisconsin, 8701 Watertown Plank Road, Milwaukee, WI
53226, USA
COMMENT approximately 100 bp to next reported sequence, GenBank Accession
Number U29918.

FEATURES
source Location/Qualifiers
1..1768
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="11"
/map="lip13-pter"
/cell_type="T-lymphocyte, cytotoxic"
/clone_lib="RPWI 8402, lambda 2001 library of R. Baer"
<1..389
/gene="AMPD3"
/number=7
390..521
/gene="AMPD3"
/number=8
522..1058
/gene="AMPD3"
/number=8
1059..1222
/gene="AMPD3"

intron
exon
intron
exon

/misc_feature
/misc_feature
/misc_feature
/misc_feature
/misc_feature
BASE COUNT 73610 a 42903 c 42981 g 69761 t 407 others
ORIGIN

/number=9
1223..>1768
/gene="AMPD3"
/number=9
379 a 459 c 496 g 434 t
BASE COUNT
ORIGIN

Query Match 74.2%; Score 17.8; DB 9; Length 1768;
Best Local Similarity 90.5%; Pred. No. 1.1e+02;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 TCTAGTGTGTCGAGGCATCT 21
||||| |||||||
Db 1401 TCTAGTGTGTCGAGGCATCT 1381

RESULT 10
AC090969/c
LOCUS AC090969 40520 bp DNA linear BCT 27-MAR-2002
DEFINITION Staphylococcus aureus clone sabac-57 strain NCTC 8325, complete
sequence.
ACCESSION AC090969
VERSION AC090969.3 GI:19747214
KEYWORDS HTG.
SOURCE Staphylococcus aureus.
ORGANISM Staphylococcus aureus
Bacteria; Firmicutes; Bacillales; Staphylococcus.
REFERENCE 1 (bases 1 to 40520)
AUTHORS Tian,R., Lin,S., Jia,H., Qian,Y., Iandolo,J., Worrell,V. and
Roe,B.A.
TITLE Staphylococcus aureus BAC Clone sabac-57
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 40520)
AUTHORS Tian,R., Lin,S., Jia,H., Qian,Y., Iandolo,J., Worrell,V. and
Roe,B.A.
TITLE Direct Submission
JOURNAL Submitted (20-MAR-2001) Department of Chemistry And Biochemistry,
The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
OK 73019, USA
REFERENCE 3 (bases 1 to 40520)
AUTHORS Tian,R., Lin,S., Jia,H., Qian,Y., Iandolo,J., Worrell,V. and
Roe,B.A.
TITLE Direct Submission
JOURNAL Submitted (27-MAR-2002) Department of Chemistry And Biochemistry,
The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
OK 73019, USA
COMMENT On Mar 27, 2002 this sequence version replaced gi:13491249.
----- Genome Center
Center: Department Of Chemistry And Biochemistry
The University Of Oklahoma
Center code:UOKNOR
-----

FEATURES
source Location/Qualifiers
1..40520
/organism="Staphylococcus aureus"
/strain="NCTC 8325"
/db_xref="taxon:1280"
/clone_lib="sabac-57"
/clone_lib="Iandolo BAC library"
13214 a 7178 c 5747 g 14381 t
BASE COUNT
ORIGIN

Query Match 74.2%; Score 17.8; DB 1; Length 40520;
Best Local Similarity 90.5%; Pred. No. 99;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 4 AGTGTGTCGAGGCATCTAGT 24
||||| |||||||
Db 36574 ATTGTTGTCGAGGCATCTAGT 36554

RESULT 11
AC113035/c

```

LOCUS AC113035 69320 bp DNA linear HTG 13-MAY-2002
 DEFINITION Mus musculus clone RP23-222J21, LOW-PASS SEQUENCE SAMPLING.
 ACCESSION AC113035
 VERSION AC113035.2 GI:20531939
 KEYWORDS HTG; HTGS_PHASE0.
 SOURCE Mus musculus.
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 69320)
 Birren, B., Linton, L., Nusbaum, C. and Lander, E.
 Unpublished
 2 (bases 1 to 69320)
 Birren, B., Linton, L., Nusbaum, C., Lander, E., Ali, A., Allen, N.,
 Anderson, S., Barna, N., Bastien, V., Boguslavsky, L., Boukhgalter, B.,
 Brown, A., Canarata, J., Campopiano, A., Chang, J., Chazaro, B.,
 Choepel, Y., Colangelo, M., Collins, S., Collamore, A., Cook, A.,
 Cooke, P., Dearellano, K., Dewar, K., Diaz, J.S., Dodge, S., Faro, S.,
 Ferreira, P., FitzHugh, W., Gage, D., Galagan, J., Gardyna, S.,
 Ginde, S., Gord, S., Goyette, M., Graham, L., Grand-Pierre, N.,
 Hagos, B., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C.,
 Kanat, A., Karatas, A., Kells, C., LaRocque, K., Lamazares, R.,
 Landers, T., Lehoczy, J., Levine, R., Liu, G., MacLean, C.,
 MacDonald, P., Major, J., Marquis, N., Matthews, C., McCarthy, M.,
 McEwan, P., McKernan, K., Meldrim, J., Meneus, L., Mihova, T.,
 Mlenga, V., Murphy, T., Naylor, J., Nguyen, C., Nicol, R., Norbu, C.,
 Norman, C.H., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J.,
 Peterson, K., Phunkhang, P., Pierre, N., Pollara, V., Raymond, C.,
 Retta, R., Rieback, M., Riley, R., Rise, C., Rogov, P., Roman, J.,
 Rosetti, M., Roy, A., Santos, R., Schauer, S., Schupbach, R., Seaman, S.,
 Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N.,
 Strauss, N., Subramanian, A., Talamas, J., Tesfaye, S., Theodore, J.,
 Topham, K., Travers, M., Travis, N., Triggilio, J., Vassiliev, H.,
 Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.J., Young, G.,
 Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.
 Direct Submission
 Submitted (25-FEB-2002) Whitehead Institute/MIT Center for Genome
 Research, 320 Charles Street, Cambridge, MA 02141, USA
 3 (bases 1 to 69320)
 Birren, B., Linton, L., Nusbaum, C., Lander, E., Ali, A., Allen, N.,
 Anderson, S., Barna, N., Bastien, V., Bloom, T., Boguslavsky, L.,
 Boukhgalter, B., Brown, A., Camarata, J., Campopiano, A., Chang, J.,
 Chazaro, B., Choepel, Y., Colangelo, M., Collins, S., Collamore, A.,
 Cook, A., Cooke, P., Dearellano, K., Dewar, K., Diaz, J.S., Dodge, S.,
 Faro, S., Ferreira, P., FitzHugh, W., Gage, D., Galagan, J., Gardyna, S.,
 Ginde, S., Gord, S., Goyette, M., Graham, L., Grand-Pierre, N.,
 Hagos, B., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C.,
 Kanat, A., Karatas, A., Kells, C., LaRocque, K., Lamazares, R.,
 Landers, T., Lehoczy, J., Levine, R., Lindblad-Toh, K., Liu, G.,
 MacLean, C., MacDonald, P., Major, J., Marquis, N., Matthews, C.,
 McCarthy, M., McEwan, P., McKernan, K., Meldrim, J., Meneus, L.,
 Mihova, T., Mlenga, V., Murphy, T., Naylor, J., Nguyen, C., Nicol, R.,
 Norbu, C., Norman, C.H., O'Connor, T., O'Donnell, P., O'Neill, D.,
 Oliver, J., Peterson, K., Phunkhang, P., Pierre, N., Pollara, V.,
 Raymond, C., Retta, R., Rieback, M., Riley, R., Rise, C., Rogov, P.,
 Roman, J., Rosetti, M., Roy, A., Santos, R., Schauer, S., Schupbach, R.,
 Seaman, S., Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N.,
 Strauss, N., Subramanian, A., Talamas, J., Tesfaye, S., Theodore, J.,
 Topham, K., Travers, M., Travis, N., Triggilio, J., Vassiliev, H.,
 Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.J., Young, G.,
 Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.
 Direct Submission
 Submitted (13-MAY-2002) Whitehead Institute/MIT Center for Genome
 Research, 320 Charles Street, Cambridge, MA 02141, USA
 On May 13, 2002 this sequence version replaced gi:18875122.
 All repeats were identified using RepeatMasker:
 Smit, A.F.A. & Green, P. (1996-1997)
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>
 ----- Genome Center
 Center: Whitehead Institute/ MIT Center for Genome Research
 Center code: WIBR
 Web site: <http://www-seq.wi.mit.edu>

Contact: sequence_submissions@genome.wi.mit.edu
 ----- Project Information
 Center project name: L23562
 Center clone name: 222_J_21

 * NOTE: This record contains 87 individual
 * sequencing reads that have not been assembled into
 * contigs. Runs of N are used to separate the reads
 * and the order in which they appear is completely
 * arbitrary. Low-pass sequence sampling is useful for
 * identifying clones that may be gene-rich and allows
 * overlap relationships among clones to be deduced.
 * However, it should not be assumed that this clone
 * will be sequenced to completion. In the event that
 * the record is updated, the accession number will
 * be preserved.
 * 1 703: contig of 703 bp in length
 * 704 803: gap of 100 bp
 * 804 1508: contig of 705 bp in length
 * 1509 1608: gap of 100 bp
 * 1609 2275: contig of 667 bp in length
 * 2276 2375: gap of 100 bp
 * 2376 3084: contig of 709 bp in length
 * 3085 3184: gap of 100 bp
 * 3185 3892: contig of 708 bp in length
 * 3893 3992: gap of 100 bp
 * 3993 4678: contig of 686 bp in length
 * 4679 4778: gap of 100 bp
 * 4779 5470: contig of 692 bp in length
 * 5471 5570: gap of 100 bp
 * 5571 6259: contig of 689 bp in length
 * 6260 6359: gap of 100 bp
 * 6360 7049: contig of 690 bp in length
 * 7050 7149: gap of 100 bp
 * 7150 7834: contig of 685 bp in length
 * 7835 7934: gap of 100 bp
 * 7935 8650: contig of 716 bp in length
 * 8651 8750: gap of 100 bp
 * 8751 9448: contig of 698 bp in length
 * 9449 9548: gap of 100 bp
 * 9549 10262: contig of 714 bp in length
 * 10263 10362: gap of 100 bp
 * 10363 11069: contig of 707 bp in length
 * 11070 11169: gap of 100 bp
 * 11170 11807: contig of 638 bp in length
 * 11808 11907: gap of 100 bp
 * 11908 12576: contig of 669 bp in length
 * 12577 12676: gap of 100 bp
 * 12677 13371: contig of 695 bp in length
 * 13372 13471: gap of 100 bp
 * 13472 14184: contig of 713 bp in length
 * 14185 14284: gap of 100 bp
 * 14285 14990: contig of 706 bp in length
 * 14991 15090: gap of 100 bp
 * 15091 15791: contig of 701 bp in length
 * 15792 15891: gap of 100 bp
 * 15892 16604: contig of 713 bp in length
 * 16605 16704: gap of 100 bp
 * 16705 17407: contig of 703 bp in length
 * 17408 17507: gap of 100 bp
 * 17508 18207: contig of 700 bp in length
 * 18208 18307: gap of 100 bp
 * 18308 19012: contig of 705 bp in length
 * 19013 19112: gap of 100 bp
 * 19113 19799: contig of 687 bp in length
 * 19800 19899: gap of 100 bp
 * 19900 20594: contig of 695 bp in length
 * 20595 20694: gap of 100 bp
 * 20695 21386: contig of 692 bp in length
 * 21387 21486: gap of 100 bp
 * 21487 22178: contig of 692 bp in length
 * 22179 22278: gap of 100 bp
 * 22279 22983: contig of 705 bp in length

```

* 22984 23083: gap of 100 bp
* 23084 23700: contig of 617 bp in length
* 23701 23800: gap of 100 bp
* 23801 24512: contig of 712 bp in length
* 24513 24612: gap of 100 bp
* 24613 25318: contig of 706 bp in length
* 25319 25418: gap of 100 bp
* 25419 26133: contig of 715 bp in length
* 26134 26233: gap of 100 bp
* 26234 26953: contig of 720 bp in length
* 26954 27053: gap of 100 bp
* 27054 27756: contig of 703 bp in length
* 27757 27856: gap of 100 bp
* 27857 28554: contig of 698 bp in length
* 28555 28654: gap of 100 bp
* 28655 29348: contig of 694 bp in length
* 29349 29448: gap of 100 bp
* 29449 30148: contig of 700 bp in length
* 30149 30248: gap of 100 bp
* 30249 30953: contig of 705 bp in length
* 30954 31053: gap of 100 bp
* 31054 31742: contig of 689 bp in length
* 31743 31842: gap of 100 bp
* 31843 32547: contig of 705 bp in length
* 32548 32647: gap of 100 bp
* 32648 33352: contig of 705 bp in length
* 33353 33452: gap of 100 bp
* 33453 34146: contig of 694 bp in length
* 34147 34246: gap of 100 bp
* 34247 34952: contig of 706 bp in length
* 34953 35052: gap of 100 bp
* 35053 35753: contig of 701 bp in length
* 35754 35853: gap of 100 bp
* 35854 36537: contig of 684 bp in length
* 36538 36637: gap of 100 bp
* 36638 37322: contig of 685 bp in length
* 37323 37422: gap of 100 bp
* 37423 38115: contig of 693 bp in length
* 38116 38215: gap of 100 bp
* 38216 38917: contig of 702 bp in length
* 38918 39017: gap of 100 bp
* 39018 39723: contig of 706 bp in length
* 39724 39823: gap of 100 bp
* 39824 40526: contig of 703 bp in length
* 40527 40626: gap of 100 bp
* 40627 41323: contig of 697 bp in length
* 41324 41423: gap of 100 bp
* 41424 42138: contig of 715 bp in length
* 42139 42238: gap of 100 bp
* 42239 42946: contig of 708 bp in length
* 42947 43046: gap of 100 bp
* 43047 43766: contig of 720 bp in length
* 43767 43866: gap of 100 bp
* 43867 44567: contig of 701 bp in length

```

```

Query Match      74.2%; Score 17.8; DB 2; Length 69320;
Best Local Similarity 90.5%; Pred. No. 97;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```

```

QY 4 AGTGTGTCGACGCATCTAGT 24
    |||||
Db 56929 AGTGTGTCGACGCATAGT 56909

```

```

RESULT 12
AC048373/c
LOCUS
DEFINITION
Homo sapiens chromosome 11 clone RP11-139C6 map 11, WORKING DRAFT
AC048373
ACCESSION
AC048373.2 GI:7798808
VERSION
KEYWORDS
HTG: HTGS_PHASE1; HTGS_DRAFT.
SOURCE
Homo sapiens.

```

ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

1 (bases 1 to 152631)

Birren, B., Linton, L., Nusbaum, C., Lander, E., Abraham, H., Allen, N.,

Anderson, S., Baldwin, J., Barna, N., Bastien, V., Beda, F.,

Boguslavsky, L., Bouckhgalter, B., Brown, A., Burkett, G.,

Campopiano, A., Castle, A., Choquel, Y., Colangelo, M., Collins, S.,

Collymore, A., Cooke, P., DeAellano, K., Dewar, K., Diaz, J.S.,

Dodge, S., Domino, M., Doyle, M., Ferreira, P., FitzHugh, W., Gage, D.,

Galagan, J., Gardyna, S., Ginde, S., Goyette, M., Graham, L.,

Grand-Pierre, N., Grant, G., Hagos, B., Hearford, A., Horton, L.,

Howland, J.C., Iliev, I., Johnson, R., Jones, C., Kann, L., Karatas, A.,

Klein, J., LaRocque, K., Lamazares, R., Landers, T., Lehoczy, J.,

Levine, R., Liu, C., Liu, G., Locke, K., Macdonald, P., Marquis, N.,

McCarthy, M., McEwan, P., McGurk, A., McKernan, K., McPheeters, R.,

Meldrum, J., Meneus, L., Mihova, T., Miranda, C., Miens, V., Morrow, J.,

Murphy, T., Naylor, J., Norman, C.H., O'Connor, T., O'Donnell, P.,

O'Neill, D., Oliver, T.M., Oliver, J., Peterson, K., Pierre, N.,

Pisani, C., Pollara, V., Raymond, C., Riley, R., Rogov, P., Rothman, D.,

Roy, A., Santos, R., Schauer, S., Severy, P., Spencer, B.,

Stange-Thomann, N., Stojanovic, N., Subramanian, A., Talamas, J.,

Tesfaye, S., Theodore, J., Tirrell, A., Travers, M., Trigglio, J.,

Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.J.,

Young, G., Zainoun, J., Zimmer, A. and Zody, M.

Direct Submission

Submitted (14-APR-2000) Whitehead Institute/MIT Center for Genome

Research, 320 Charles Street, Cambridge, MA 02141, USA

On May 13, 2000 this sequence version replaced gi:7549705.

All repeats were identified using RepeatMasker:

Smit, A.F.A. & Green, P. (1996-1997)

http://ftp.genome.washington.edu/RM/RepeatMasker.html

----- Genome Center

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WIBR

Web site: http://www-seq.wi.mit.edu

Contact: sequence_submissions@genome.wi.mit.edu

----- Project Information

Center project name: L8701

Center clone name: 139_C_6

----- Summary Statistics

Sequencing vector: M13; M77815; 100% of reads

Chemistry: Dye-terminator Big Dye; 100% of reads

Assembly program: Phrap; version 0.960731

Consensus quality: 145913 bases at least Q40

Consensus quality: 149504 bases at least Q30

Consensus quality: 150798 bases at least Q20

Insert size: 151000; agarose-1p

Quality coverage: 151631; sum-of-contigs

Quality coverage: 4.8 in Q20 bases; agarose-1p

Quality coverage: 4.8 in Q20 bases; sum-of-contigs

* NOTE: This is a 'working draft' sequence. It currently

* consists of 11 contigs. The true order of the pieces

* is not known and their order in this sequence record is

* arbitrary. Gaps between the contigs are represented as

* runs of N, but the exact sizes of the gaps are unknown.

* This record will be updated with the finished sequence

* as soon as it is available and the accession number will

* be preserved.

* 1 1150: contig of 1150 bp in length

* 1151 1250: gap of 100 bp

* 1251 6488: contig of 5238 bp in length

* 6489 6588: gap of 100 bp

* 6589 11741: contig of 5153 bp in length

* 11742 11841: gap of 100 bp

* 11842 16228: contig of 4387 bp in length

* 16229 16328: gap of 100 bp

* 16329 28016: contig of 11688 bp in length

```

* 28017 28116: gap of 100 bp
* 28117 45296: contig of 17180 bp in length
* 45297 45396: gap of 100 bp
* 45397 59567: contig of 14171 bp in length
* 59568 59667: gap of 100 bp
* 59668 76149: contig of 16482 bp in length
* 76150 76249: gap of 100 bp
* 76250 92365: contig of 16116 bp in length
* 92366 92465: gap of 100 bp
* 92466 119998: contig of 27533 bp in length
* 119999 120098: gap of 100 bp
* 120099 152631: contig of 32533 bp in length.

```

FEATURES

source

```

Location/Qualifiers
1. .152631
  /organism="Homo sapiens"
  /db_xref="taxon:9606"
  /chromosome="11"
  /map="11"
  /clone="RP11-13906"
  /clone_lib="RPC1-11 Human Male BAC"

```

```

misc_feature
1. .1150
  /note="assembly_fragment"

```

```

misc_feature
1251. 6488
  /note="assembly_fragment"

```

```

misc_feature
6589. .11741
  /note="assembly_fragment"

```

```

misc_feature
11842. .16228
  /note="assembly_fragment"

```

```

misc_feature
clone_end.SP6
vector_side:left
16329. .28016
  /note="assembly_fragment"

```

```

misc_feature
clone_end.T7
vector_side:right
28117. .45296
  /note="assembly_fragment"

```

```

misc_feature
45397. .59567
  /note="assembly_fragment"

```

```

misc_feature
59668. .76149
  /note="assembly_fragment"

```

```

misc_feature
76250. 92365
  /note="assembly_fragment"

```

```

misc_feature
92466. .119998
  /note="assembly_fragment"

```

```

misc_feature
120099. .152631
  /note="assembly_fragment"

```

```

BASE COUNT 49516 a 26157 c 27027 g 48926 t 1005 others
ORIGIN

```

```

Query Match 74.28; Score 17.8; DB 2; Length 152631;
Best Local Similarity 90.5%; Pred. No. 94;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```

```

QY 1 TCTAGTGTGTCGAGGCATCT 21
||||| ||||| ||||| |||||

```

```

Db 28560 TCTAGTGTGTCGAGGCATCT 28540

```

RESULT 13

AC021914

```

LOCUS AC021914 156507 bp DNA linear PRI 03-NOV-2001
DEFINITION Homo sapiens chromosome 11, clone RP11-68C8, complete sequence.
ACCESSION AC021914
VERSION AC021914.7 GI:16304268
KEYWORDS HTG.
SOURCE Homo sapiens.

```

ORGANISM

```

Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 156507)

```

REFERENCE

AUTHORS

```

Birren,B., Linton,L., Nusbaum,C. and Lander,E.

```

TITLE

```

Homo sapiens chromosome 11, clone RP11-68C8

```

JOURNAL

```

Unpublished

```

REFERENCE

AUTHORS

```

2 (bases 1 to 156507)
Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
Anderson,S., Baldwin,J., Barna,N., Beckerly,R., Bada,F.,
Boguslavsky,L., Boukhgalter,B., Brown,A., Burkett,G., Castle,A.,
Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cooke,P.,
DeArelano,K., Dewar,K., Domino,M., Doyle,M., Fenestor,J.,
Ferreira,P., FitzHugh,W., Forrest,C., Gage,D., Galagan,J.,
Gardyna,S., Grant,G., Hagos,B., Headford,A., Horton,L.,
Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,
Landers,T., Lehoczy,J., Levine,R., Lieu,C., Liu,S., Locke,K.,
Macdonald,P., Marquis,N., McEwan,P., McGurk,A., McKernan,K.,
McPheeters,R., Meldrum,J., Meneus,L., Morrow,J., Naylor,J.,
Norman,C.H., O'Connor,T., O'Donnell,P., Olivari,T.M., Peterson,K.,
Pierre,N., Pisani,C., Pollara,V., Raymond,C., Riley,R., Rothman,D.,
Roy,A., Santos,R., Severy,P., Spencer,B., Stange-Thomann,N.,
Stojanovic,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
Tirrell,A., Vassiliev,H., Viel,R., Vo,A., Wu,X., Wyman,D., Ye,W.J.,
Zimmer,A. and Zody,M.
Direct Submission
Submitted (22-JAN-2000) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
3 (bases 1 to 156507)
Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
Anderson,S., Barna,N., Bastien,V., Boguslavsky,L., Boukhgalter,B.,
Brown,A., Camarata,J., Campiano,A., Chang,J., Chazaro,B.,
Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cook,A.,
Cooke,P., DeArelano,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S.,
Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S.,
Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
Hagos,B., Headford,A., Horton,L., Hulme,W., Iliev,I., Johnson,R.,
Jones,C., Kamat,A., Karatas,A., Kells,C., LaRocque,K.,
Lamazares,R., Landers,T., Lehoczy,J., Levine,R., Liu,G.,
MacLean,C., Macdonald,P., Major,J., Marquis,N., Matthews,C.,
McCarthy,M., McEwan,P., McKernan,K., McPheeters,R., Meldrum,J.,
Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C.,
Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D.,
Olivari,T., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V.,
Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P.,
Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schuback,R.,
Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
Topham,K., Travers,M., Travis,N., Triggillo,J., Vassiliev,H.,
Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
Direct Submission
Submitted (03-NOV-2001) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Oct 21, 2001 this sequence version replaced gi:16259015.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L4204
Center clone name: 68_C8
-----
Location/Qualifiers
1. .156507
  /organism="Homo sapiens"
  /db_xref="taxon:9606"
  /chromosome="11"
  /map="11"
  /clone="RP11-68C8"
  /clone_lib="RPC1-11 Human Male BAC"
  complement(132..1126)
  /rpt_family="LIMC3"
  1129..1388
  /rpt_family="LPR26B"
  1433..1588

```

TITLE

JOURNAL

COMMENT

COMMENT

FEATURES

source

```

repeat_region
repeat_region
repeat_region

```



```

repeat_region /rpt_family="LTR67"
repeat_region complement(1794..2350)
repeat_region /rpt_family="MLT1F1"
repeat_region 3141..3440
repeat_region /rpt_family="AluY"
repeat_region 5555..5749
repeat_region /rpt_family="MIR3"
repeat_region 5896..6215
repeat_region /rpt_family="AluSc"
repeat_region 6683..6721
repeat_region /rpt_family="AT_rich"
repeat_region complement(6896..8100)
repeat_region /rpt_family="LIME1"
repeat_region 8101..9687
repeat_region /rpt_family="MERS2A"
repeat_region complement(9688..10187)
repeat_region /rpt_family="LIME1"
repeat_region complement(10188..10601)
repeat_region /rpt_family="LIPAL6"
repeat_region complement(10602..10960)
repeat_region /rpt_family="LIME1"
repeat_region complement(11102..11204)
repeat_region /rpt_family="L2"
repeat_region 11218..11462
repeat_region /rpt_family="MIR"
repeat_region 11553..11644
repeat_region /rpt_family="MIR"
repeat_region 11664..11701
repeat_region /rpt_family="L2"
repeat_region complement(14831..14874)
repeat_region /rpt_family="L2"
repeat_region 15458..15970
repeat_region /rpt_family="MER34"
repeat_region complement(16175..16703)
repeat_region /rpt_family="L3"
repeat_region 16721..16864
repeat_region /rpt_family="FLAM_C"
repeat_region complement(16867..17095)
repeat_region /rpt_family="L3"
repeat_region 17119..17522
repeat_region /rpt_family="L2"
repeat_region complement(17896..17983)
repeat_region /rpt_family="L3"
repeat_region 18245..18266
repeat_region /rpt_family="AT_rich"
repeat_region 18587..18612
repeat_region /rpt_family="AT_rich"
repeat_region 18683..18921
repeat_region /rpt_family="MIR"
repeat_region 18976..19004
repeat_region /rpt_family="(TG)n"
repeat_region complement(20451..20507)
repeat_region /rpt_family="L2"
repeat_region 20944..20985
repeat_region /rpt_family="(TG)n"
repeat_region complement(22094..22386)
repeat_region /rpt_family="AluSg"
repeat_region 22673..22709
repeat_region /rpt_family="AT_rich"
repeat_region complement(23646..23730)
repeat_region /rpt_family="L2"
repeat_region 23737..23890
repeat_region /rpt_family="L2"
repeat_region complement(24204..24431)
repeat_region /rpt_family="MLT1L"
repeat_region 24464..24727
repeat_region /rpt_family="L2"
repeat_region 25491..26366
repeat_region /rpt_family="L2"
repeat_region complement(26683..26980)
repeat_region /rpt_family="AluY"
repeat_region 27086..27106
repeat_region /rpt_family="AT_rich"

```

```

repeat_region complement(29065..29349)
repeat_region /rpt_family="AluY"
repeat_region complement(29980..30201)
repeat_region /rpt_family="MIR"
repeat_region complement(30234..30327)
repeat_region /rpt_family="MERSA"
repeat_region 30518..30740
repeat_region /rpt_family="MIR"
repeat_region complement(30753..31085)
repeat_region /rpt_family="L3"
repeat_region 31411..31476
repeat_region /rpt_family="L2"
repeat_region complement(31861..31991)
repeat_region /rpt_family="L2"
repeat_region complement(32022..32102)
repeat_region /rpt_family="LIME4A"
repeat_region 32409..32469
repeat_region /rpt_family="L2"
repeat_region 32589..33390
repeat_region /rpt_family="Tigger3b"
repeat_region complement(33391..33606)
repeat_region /rpt_family="MER30"
repeat_region 33607..33988
repeat_region /rpt_family="Tigger3b"
repeat_region complement(34784..34848)
repeat_region /rpt_family="MIR3"
repeat_region complement(34875..34947)
repeat_region /rpt_family="LIME2"
repeat_region complement(34957..35356)
repeat_region /rpt_family="LIME2"
repeat_region complement(35352..35689)
repeat_region /rpt_family="LIME2"
repeat_region complement(35706..37242)
repeat_region /rpt_family="LIME2"
repeat_region 37965..38151
repeat_region /rpt_family="LIME4a"
repeat_region 38241..38361
repeat_region /rpt_family="HAL1"
repeat_region 39513..39804
repeat_region /rpt_family="AluSx"
repeat_region 40701..40730

```

Query Match 74.2%; Score 17.8; DB 9; Length 156507;
 Best Local Similarity 90.5%; Pred. No. 94;
 Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 TCTAGTCTGTCGAGCATCT 21
 ||||| ||||| ||||| |||||
 Db 57570 TCTAGTCTGTCGAGCATCT 57590

RESULT 14
 LOCUS AP002378/c 177429 bp DNA linear PRI 26-APR-2001
 DEFINITION Homo sapiens genomic DNA, chromosome 11q, clone:RP11-351D3,
 complete sequences.
 ACCESSION AP002378
 VERSION AP002378.3 GI:13810520
 KEYWORDS HTG.
 SOURCE Homo sapiens
 ORGANISM Homo sapiens DNA, clone:RP11-351D3.
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (sites)
 AUTHORS Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
 Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
 TITLE Homo sapiens genomic DNA
 JOURNAL Published Only in Database
 REFERENCE 2 (bases 1 to 177429)
 AUTHORS Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
 Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
 TITLE Direct Submission
 JOURNAL Submitted (29-MAY-2000) Masahira Hattori, The Institute of Physical

and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
1-7-22 Suehiro-chou, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
(E-mail: hattori@gsc.riken.go.jp, URL: http://hgp.gsc.riken.go.jp/,
Tel: 81-45-503-9111, Fax: 81-45-503-9170)

On Apr 26, 2001 this sequence version replaced gi:11136462.

COMMENT FEATURES

source

1..177429
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="11"
/map="11q"
/clone="RP11-351D3"
BASE COUNT 57127 a 30586 c 31846 g 57870 t
ORIGIN

Query Match 74.2%; Score 17.8; DB 9; Length 177429;

Best Local Similarity 90.5%; Pred. No. 94;

Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 TCTAGTGTGTCGAGGCATCT 21

||||| |||||||

Db 102035 TCTAGTGGAGTCGAGGCATCT 102015

RESULT 15

AP000760/c

LOCUS

DEFINITION

AP000760

complete sequences.

AP000760.4 GI:13810525

HTG.

AP000760

HTG.

AP000760

HTG.

HTG.

HTG.

HTG.

HTG.

HTG.

HTG.

HTG.

HTG.

HTG.

HTG.

HTG.

HTG.

HTG.

HTG.

HTG.

HTG.

HTG.

HTG.

HTG.

HTG.

HTG.

HTG.

HTG.

HTG.

HTG.

HTG.

HTG.

HTG.

HTG.

HTG.

HTG.

HTG.

HTG.

HTG.

HTG.

HTG.

HTG.

HTG.

HTG.

HTG.

HTG.

HTG.

HTG.

HTG.

HTG.

Query Match 74.2%; Score 17.8; DB 9; Length 177488;

Best Local Similarity 90.5%; Pred. No. 94;

Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 TCTAGTGTGTCGAGGCATCT 21

||||| |||||||

Db 3629 TCTAGTGGAGTCGAGGCATCT 3609

Search completed: January 3, 2003, 23:57:17

Job time : 371.151 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 3, 2003, 22:52:02 ; Search time 40.1262 Seconds
(without alignments)
1346.950 Million cell updates/sec

Title: US-09-787-562-11

Perfect score: 24

Sequence: 1 tctagtctgtgcaggcatctagt 24

Scoring table: IDENTITY_NUC

Gapop 10_0 , Gapext 1.0

Searched: 2185239 seqs, 1125999159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N_Genesd_101002.*

1: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1980.DAT.*
2: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1981.DAT.*
3: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1982.DAT.*
4: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1983.DAT.*
5: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1984.DAT.*
6: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1985.DAT.*
7: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1986.DAT.*
8: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1987.DAT.*
9: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1988.DAT.*
10: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1989.DAT.*
11: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1990.DAT.*
12: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1991.DAT.*
13: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1992.DAT.*
14: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1993.DAT.*
15: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1994.DAT.*
16: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1995.DAT.*
17: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1996.DAT.*
18: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1997.DAT.*
19: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1998.DAT.*
20: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1999.DAT.*
21: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2000.DAT.*
22: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT.*
23: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT.*
24: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	24	100.0	24	21	AA12003
2	24	100.0	237	21	AA12001
3	17.8	74.2	2817	18	AAV74606
C 4	17.4	72.5	72	20	AAZ11441
C 5	17.4	72.5	72	21	AA111998
C 6	17.4	72.5	72	21	AA112024
C 7	17.4	72.5	72	21	AA112025
C 8	16.6	69.2	39	8	AAV70749
C 9	16.6	69.2	111	22	AA125330
					Murine HIF-1 space
					Murine PGK HRE der
					Staphylococcus aur
					Mutant HRE-contain
					Murine PGK HRE-con
					Murine PGK HRE HIF
					Murine PGK HRE-con
					Sequence of portio
					Human breast cance

C 10	16.6	69.2	115	22	AA116487	Human breast cance
C 11	16.6	69.2	192	23	ABV14349	Human prostate exp
C 12	16.6	69.2	202	21	AA109061	Human prostate exp
C 13	16.6	69.2	450	23	ABV44262	Human prostate exp
C 14	16.6	69.2	479	23	ABV35433	Human prostate exp
C 15	16.6	69.2	726	22	AA10807	Human Janus kinase
C 16	16.6	69.2	749	22	ABA09516	Human endozepine-r
C 17	16.6	69.2	749	22	AAH99188	Human protein enco
C 18	16.6	69.2	1498	22	AAK78096	Human immune/haema
C 19	16.6	69.2	1593	22	AAF81568	Human endozepine-1
C 20	16.6	69.2	1677	24	AA143759	Human NOV2 gene se
C 21	16.6	69.2	1727	21	AAA47432	Sequence encoding
C 22	16.6	69.2	1747	24	ABN88960	Human NOV1 encodin
C 23	16.6	69.2	1930	22	AA10807	Human SEC8 nucleic
C 24	16.6	69.2	2056	22	AA101220	DNA encoding human
C 25	16.6	69.2	2056	22	AA101220	Human SEC12 nucle
C 26	16.6	69.2	2510	22	AAK51825	Human polynucleoti
C 27	16.6	69.2	3807	17	AA130862	Protein tyrosine k
C 28	16.6	69.2	3887	24	AA143754	Human NOV1a gene s
C 29	16.6	69.2	3920	24	AA143755	Human NOV1b gene s
C 30	16.6	69.2	3920	24	AA143756	Human NOV1c gene s
C 31	16.6	69.2	3920	24	AA143757	Human NOV1d gene s
C 32	16.6	69.2	3920	24	AA143758	Human NOV1e gene s
C 33	16.6	69.2	6781	19	AAV45824	Maize phosphoenolp
C 34	16.6	69.2	7559	20	AAV82011	Human immune/haema
C 35	16.6	69.2	26059	22	AAK69104	Bacillus clausil g
C 36	16.2	67.5	557	24	ABK77504	Staphylococcus aur
C 37	16.2	67.5	902	18	AAV75089	HSP47 gene fragmen
C 38	16.2	67.5	1659	20	AA227698	N. meningitidis pa
C 39	16.2	67.5	1698	21	AA181571	HRE-containing enh
C 40	16.6	66.7	72	21	AA111440	Murine PGK HRE HIF
C 41	16.6	66.7	72	21	AA111440	HRE element contai
C 42	16.6	66.7	80	20	AA111431	EIAV U3 enhancer r
C 43	16.6	66.7	100	21	AA112060	Human foetal liver
C 44	16.6	66.7	114	21	AA112061	
C 45	16.6	66.7	159	22	ABA71452	

ALIGNMENTS

RESULT 1

AA12003

ID AAA12003 standard; DNA: 24 BP.

XX

AC AAA12003;

XX

DT 14-AUG-2000 (first entry)

XX

DE Murine HIF-1 spacer DNA #2.

XX

DE HRE; hypoxia response element; hypoxia-inducible factor; HIF; vasotropic;
KW cardiant; cycostatic; antiarthritic; gene therapy; ischaemia; arthritis;
KW cardiovascular disease; peripheral arterial disease; cancer; murine; ds.

XX

OS Mus sp.

XX

PN WO200017371-A1.

XX

PD 30-MAR-2000.

XX

PF 22-SEP-1999; 99WO-GB03181.

XX

PR 23-SEP-1998; 98WO-GB02885.

PR 28-JAN-1999; 99GB-0001906.

PR 16-FEB-1999; 99GB-0003538.

PA (OXFO-) OXFORD BIOMEDICA UK LTD.

XX

PI Binley KM, Naylor S;

XX

XX WPI; 2000-283595/24.

XX

PT Novel polynucleotide constructs comprising at least two repeats of a
 PT hypoxia response element useful for driving expression of nucleic acids
 PT of interest in a cell under hypoxic conditions -
 XX
 PS Disclosure; Page 10; 155pp; English.
 XX
 CC This invention describes novel polynucleotide comprising at least 2
 CC repeats of a hypoxia response element (HRE), where the hypoxia-inducible
 CC factor (HIF) consensus binding sites within each of the 2 repeats are
 CC separated by a spacer of at least 20 contiguous nucleotides. The products
 CC of the invention have vasotropic, cardiant, cytostatic and antiarthritic
 CC activity and can be used for gene therapy. The polynucleotides are useful
 CC for delivering nucleic acids of interest to mammalian cells. Lentiviral
 CC vectors are responsive to hypoxic agents and to agents that mimic
 CC hypoxia. This regulation can be harnessed in vitro to enhance the
 CC production of the vector and can be used in vivo to regulate gene
 CC expression in response to a physiological signal. The vectors have
 CC utility in disease, where ischaemia, including hypoxia, is a feature,
 CC e.g. cardiovascular disease, peripheral arterial disease, cancer and
 CC arthritis. The novel regulatory construct is capable of driving very high
 CC levels of transcription under conditions of hypoxia whilst providing only
 CC low basal levels of transcription under normal oxygen conditions. The
 CC polynucleotide construct targets cells within a tumor mass that are under
 CC conditions of hypoxia without affecting normal surrounding tissue. This
 CC sequence represents a murine HIF-1 DNA spacer which is used in the method
 CC of the invention.

SQ Sequence 24 BP; 4 A; 5 C; 7 G; 8 T; 0 other;

Query Match 100.0%; Score 24; DB 21; Length 24;
 Best Local Similarity 100.0%; Pred. No. 0.013;
 Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TCTAGTGTCTGTCAGGCATCTAGT 24
 |||||
 Db 1 TCTAGTGTCTGTCAGGCATCTAGT 24

RESULT 2
 AAAL2001
 ID AAAL2001 standard; DNA; 237 BP.
 AC AAAL2001;
 XX
 DT 14-AUG-2000 (first entry)
 XX
 DE Murine PGK HRE derived promoter Obhrell DNA.
 XX
 KW HRE; hypoxia response element; hypoxia-inducible factor; HIF; vasotropic;
 KW cardiant; cytostatic; antiarthritic; gene therapy; ischaemia; arthritis;
 KW cardiovascular disease; peripheral arterial disease; cancer;
 KW phosphoglycerate kinase; PGK; murine; promoter; Obhrell; ds.
 XX
 OS Mus sp.
 XX
 PN WO200017371-A1.
 XX
 PD 30-MAR-2000.
 XX
 PF 22-SEP-1999; 99WO-GB03181.
 XX
 PR 23-SEP-1998; 98WO-GB02885.
 PR 28-JAN-1999; 99GB-0001906.
 PR 16-FEB-1999; 99GB-0003538.
 XX
 PA (OXFO-) OXFORD BIOMEDICA UK LTD.
 XX
 PI Binley KM, Naylor S;
 XX
 DR WPI; 2000-283595/24.
 XX
 XX Novel polynucleotide constructs comprising at least two repeats of a
 PT hypoxia response element useful for driving expression of nucleic acids

PT of interest in a cell under hypoxic conditions -
 XX
 PS Example 1; Page 68; 155pp; English.
 XX
 CC This invention describes novel polynucleotide comprising at least 2
 CC repeats of a hypoxia response element (HRE), where the hypoxia-inducible
 CC factor (HIF) consensus binding sites within each of the 2 repeats are
 CC separated by a spacer of at least 20 contiguous nucleotides. The products
 CC of the invention have vasotropic, cardiant, cytostatic and antiarthritic
 CC activity and can be used for gene therapy. The polynucleotides are useful
 CC for delivering nucleic acids of interest to mammalian cells. Lentiviral
 CC vectors are responsive to hypoxic agents and to agents that mimic
 CC hypoxia. This regulation can be harnessed in vitro to enhance the
 CC production of the vector and can be used in vivo to regulate gene
 CC expression in response to a physiological signal. The vectors have
 CC utility in disease, where ischaemia, including hypoxia, is a feature,
 CC e.g. cardiovascular disease, peripheral arterial disease, cancer and
 CC arthritis. The novel regulatory construct is capable of driving very high
 CC levels of transcription under conditions of hypoxia whilst providing only
 CC low basal levels of transcription under normal oxygen conditions. The
 CC polynucleotide construct targets cells within a tumor mass that are under
 CC conditions of hypoxia without affecting normal surrounding tissue. This
 CC sequence represents a murine phosphoglycerate kinase (PGK) HRE derived
 CC promoter Obhrell which is described in the method of the invention.

XX Sequence 237 BP; 43 A; 82 C; 56 G; 56 T; 0 other;

Query Match 100.0%; Score 24; DB 21; Length 237;
 Best Local Similarity 100.0%; Pred. No. 0.019;
 Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TCTAGTGTCTGTCAGGCATCTAGT 24
 |||||
 Db 25 TCTAGTGTCTGTCAGGCATCTAGT 48

RESULT 3
 AAV74606
 ID AAV74606 standard; DNA; 2817 BP.
 AC AAV74606;
 XX
 DT 16-MAR-1999 (first entry)
 XX
 DE Staphylococcus aureus contig SEQ ID #295.

Computer readable medium; vaccine; S aureus infection; immunodetection;
 cellulitis; eyelid infection; food poisoning; osteomyelitis; therapy;
 skin infection; surgical wound infection; scalded skin syndrome;
 toxic shock syndrome; ds.
 Staphylococcus aureus.

Key Location/Qualifiers
 misc_feature 601..860
 /tag= a
 /note= "these bases represent a line of missing text in
 the sequence listing in the specification. They
 are included to maintain the nucleotide numbering
 given in the specification for this DNA sequence"
 misc_feature 2401..2460
 /tag= b
 /note= "these bases represent a line of missing text in
 the sequence listing in the specification. They
 are included to maintain the nucleotide numbering
 given in the specification for this DNA sequence"

EP786519-A2.
 30-JUL-1997.
 07-JAN-1997; 97EP-0100117.
 XX

PR 05-JAN-1996; 96US-0009861.
 XX (HUMA-) HUMAN GENOME SCI INC.
 PA Barash SC, Choi GH, Dillon PJ, Fannon MR, Kunsch CA;
 PI Rosen CA;
 PT WPI; 1997-374922/35.
 DR Polynucleotide(s) and proteins derived from *Staphylococcus aureus*
 XX stored on computer readable medium and used in the production of
 PT anti-S.aureus vaccines
 PS Claim 1; Page 1120-1122; 3271pp; English.
 XX This sequence represents one of 5191 *Staphylococcus aureus* DNA sequences
 CC of the invention. The DNA sequences are recorded on a computer readable
 CC medium, preferably selected from a floppy or hard disk, random access
 CC memory (RAM), read-only memory (ROM) or CD-ROM. Homology searches using
 CC the S.aureus DNA sequences allows putative functions to be assigned so
 CC that protein-encoding or regulatory regions of commercial, therapeutic or
 CC industrial importance can be obtained. Specifically, sequences which are
 CC likely to encode antigens have been identified and these polypeptides can
 CC be used in a vaccine composition against S.aureus infection. The
 CC polypeptides can also be used in a kit for the immunodetection of
 CC S.aureus in a sample. S.aureus is implicated in numerous human diseases,
 CC including cellulitis, eyelid infections, food poisoning, osteomyelitis,
 CC skin and surgical wound infections, scalded skin syndrome, toxic shock
 CC syndrome, etc. Organisms transformed with the DNA sequences can be used
 CC for recombinant production of the polypeptides. The new DNA sequences
 CC (and their fragments) are useful as primers or probes for isolating
 CC homologues of any of the S.aureus DNA sequences contained on the
 CC computer readable medium.
 XX
 SQ Sequence 2817 BP; 912 A; 330 C; 446 G; 999 T; 130 other;
 Query Match 74.2%; Score 17.8; DB 18; Length 2817;
 Best Local Similarity 90.5%; Pred. No. 36;
 Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
 QY 4 AGTGTGCTGCAGGCATCTAGT 24
 | ||| ||||| ||||| |||||
 Db 1105 ATTGTGTGCTGCAGGCATCTAGT 1125
 RESULT 4
 AAZ11441/C
 ID AAZ11441 standard; DNA; 72 BP.
 AC AAZ11441;
 XX 26-OCT-1999 (first entry)
 DT Mutant HRE-containing enhancer MUT PGK18+.
 DE Retroviral vector; functional splice donor site; hybrid viral vector;
 KW functional splice acceptor site; in vivo gene delivery; therapeutic;
 KW lentiviral vector; modified hematopoietic stem cell; MHC; tumour;
 KW ischemia; hypoxia response element; HRE; hypoxia; promoter; ds.
 OS Synthetic.
 OS Mus sp.
 XX WO9915684-A2.
 PN 01-APR-1999.
 PD 23-SEP-1998; 98WO-GB02885.
 PF 25-SEP-1997; 97GB-0020465.
 PR 23-SEP-1997; 97GB-0020216.
 XX (OXFO-) OXFORD BIOMEDICA UK LTD.

XX Bebbington C, Binley KM, Lewis C, Naylor S;
 XX WPI; 1999-263482/22.
 DR New retroviral vectors, for, e.g. delivering nucleotide sequences to
 XX solid tumor sites
 PT Example 5 (page 77); Fig 13 (Page 16/43); 288pp; English.
 PS The invention relates to a retroviral vector (RVV) comprising a
 CC functional splice donor site (FSDS) and a functional splice acceptor site
 CC (FSAS) where: (i) the FSDS and the FSAS flank a first nucleotide sequence
 CC of interest (NOI); (ii) the FSDS is upstream of the FSAS; (iii) the RVV
 CC is derived from a retroviral pro-vector; (iv) the retroviral pro-vector
 CC comprises a first nucleotide sequence (NS) capable of yielding the FSDS
 CC and a second NS capable of yielding the FSAS; and (v) the first NS is
 CC downstream of the second NS, such that the RVV is formed as a result of
 CC reverse transcription of the retroviral pro-vector. A hybrid viral vector
 CC (VV) system for in vivo gene delivery, which system comprises a primary
 CC a first target cell and of expressing the secondary VV, which secondary
 CC vector is capable of transducing a secondary target cell, where the
 CC primary vector is obtainable from or is based on a adenoviral vector and
 CC the secondary VV is obtainable from or is based on a RVV preferably a
 CC lentiviral vector (LVV) is also provided. The systems can be used for
 CC delivering NOIs to one or more target sites. The NOIs may encode
 CC therapeutic or diagnostic agents. The methods are used particularly for
 CC producing modified hematopoietic stem cells (MHCs) to deliver NOIs to
 CC sites such as solid tumours which are characterised by ischemia, such as
 CC hypoxia or low glucose concentration. The system permits the stable
 CC expression of NOIs in targeted cells, e.g. rapidly dividing cells. The
 CC present sequence represents a mutant HRE-containing enhancer sequence
 CC derived from PGK gene.
 XX
 SQ Sequence 72 BP; 16 A; 25 C; 12 G; 19 T; 0 other;
 Query Match 72.5%; Score 17.4; DB 20; Length 72;
 Best Local Similarity 94.7%; Pred. No. 34;
 Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 QY 1 TCTAGTGTCTGTCAGGCAT 19
 | ||||| ||||| |||||
 Db 29 TCTAGTGTCTGTCAGGCAT 11
 RESULT 5
 AAAL1998/C
 ID AAAL1998 standard; DNA; 72 BP.
 AC AAAL1998;
 XX 14-AUG-2000 (first entry)
 DT Murine PGK HRE-containing enhancer element WT PGK18+++ DNA.
 DE HRE; hypoxia response element; hypoxia-inducible factor; HIF; vasotropic;
 KW cardiant; cytostatic; antiarthritic; gene therapy; ischaemia; arthritis;
 KW cardiovascular disease; peripheral arterial disease; cancer;
 KW phosphoglycerate kinase; PGK; murine; enhancer; ds.
 OS Mus sp.
 OS WO200017371-A1.
 PN 30-MAR-2000.
 PD 22-SEP-1999; 99WO-GB03181.
 PF 23-SEP-1998; 98WO-GB02885.
 PR 28-JAN-1999; 99GB-0001906.
 PR 16-FEB-1999; 99GB-0003538.
 XX

Mon Jan 6 15:20:20 2003

PA (OXFO-) OXFORD BIOMEDICA UK LTD.
 XX Binley KM, Naylor S;
 PI WPI; 2000-283595/24.
 XX Novel polynucleotide constructs comprising at least two repeats of a
 PT hypoxia response element useful for driving expression of nucleic acids
 PT of interest in a cell under hypoxic conditions -
 XX Example 6; Page 79; 155pp; English.
 PS This invention describes novel polynucleotide comprising at least 2
 XX repeats of a hypoxia response element (HRE), where the hypoxia-inducible
 CC factor (HIF) consensus binding sites within each of the 2 repeats are
 CC separated by a spacer of at least 20 contiguous nucleotides. The products
 CC of the invention have vasotropic, cardiant, cytostatic and antiarthritic
 CC activity and can be used for gene therapy. The polynucleotides are useful
 CC for delivering nucleic acids of interest to mammalian cells. Lentiviral
 CC vectors are responsive to hypoxic agents and to agents that mimic
 CC hypoxia. This regulation can be harnessed in vitro to enhance the
 CC production of the vector and can be used in vivo to regulate gene
 CC expression in response to a physiological signal. The vectors have
 CC utility in disease, where ischaemia, including hypoxia, is a feature,
 CC e.g. cardiovascular disease, peripheral arterial disease, cancer and
 CC arthritis. The novel regulatory construct is capable of driving very high
 CC levels of transcription under conditions of hypoxia whilst providing only
 CC low basal levels of transcription under normal oxygen conditions. The
 CC polynucleotide construct targets cells within a tumor mass that are under
 CC conditions of hypoxia without affecting normal surrounding tissue. This
 CC sequence represents a murine phosphoglycerate kinase (PGK) HRE derived
 CC enhancer which is described in the method of the invention.
 XX Sequence 72 BP; 16 A; 25 C; 12 G; 19 T; 0 other;
 SQ Query Match 72.5%; Score 17.4; DB 21; Length 72;
 Best Local Similarity 94.7%; Pred. No. 34;
 Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 QY 1 TCTAGTGTCTGTCGAGGCAT 19
 |||||
 Db 29 TCTAGTGTCTGTCGAGGAAT 11
 RESULT 6
 AAA12024/c
 ID AAA12024 standard; DNA; 72 BP.
 XX AC AAA12024;
 XX 14-AUG-2000 (first entry)
 XX Murine PGK HRE HIF-1 binding site mutant DNA.
 DE HRE; hypoxia response element; hypoxia-inducible factor; HIF; vasotropic;
 KW cardiant; cytostatic; antiarthritic; gene therapy; ischaemia; arthritis;
 KW cardiovascular disease; peripheral arterial disease; cancer;
 KW phosphoglycerate kinase; PGK; murine; ds.
 XX Mus sp.
 OS WO200017371-A1.
 XX 30-MAR-2000.
 XX 22-SEP-1999; 99WO-GB03181.
 XX 23-SEP-1998; 98WO-GB02885.
 XX 28-JAN-1999; 99GB-0001906.
 XX 16-FEB-1999; 99GB-0003538.
 XX (OXFO-) OXFORD BIOMEDICA UK LTD.
 PA Binley KM, Naylor S;
 PI

PI Binley KM, Naylor S;
 XX WPI; 2000-283595/24.
 XX Novel polynucleotide constructs comprising at least two repeats of a
 PT hypoxia response element useful for driving expression of nucleic acids
 PT of interest in a cell under hypoxic conditions -
 XX Example 4; Page 76; 155pp; English.
 PS This invention describes novel polynucleotide comprising at least 2
 XX repeats of a hypoxia response element (HRE), where the hypoxia-inducible
 CC factor (HIF) consensus binding sites within each of the 2 repeats are
 CC separated by a spacer of at least 20 contiguous nucleotides. The products
 CC of the invention have vasotropic, cardiant, cytostatic and antiarthritic
 CC activity and can be used for gene therapy. The polynucleotides are useful
 CC for delivering nucleic acids of interest to mammalian cells. Lentiviral
 CC vectors are responsive to hypoxic agents and to agents that mimic
 CC hypoxia. This regulation can be harnessed in vitro to enhance the
 CC production of the vector and can be used in vivo to regulate gene
 CC expression in response to a physiological signal. The vectors have
 CC utility in disease, where ischaemia, including hypoxia, is a feature,
 CC e.g. cardiovascular disease, peripheral arterial disease, cancer and
 CC arthritis. The novel regulatory construct is capable of driving very high
 CC levels of transcription under conditions of hypoxia whilst providing only
 CC low basal levels of transcription under normal oxygen conditions. The
 CC polynucleotide construct targets cells within a tumor mass that are under
 CC conditions of hypoxia without affecting normal surrounding tissue. This
 CC sequence represents a murine wild type phosphoglycerate kinase (PGK) HRE
 CC HIF-1 binding site which is described in the method of the invention.
 XX Sequence 72 BP; 16 A; 25 C; 12 G; 19 T; 0 other;
 SQ Query Match 72.5%; Score 17.4; DB 21; Length 72;
 Best Local Similarity 94.7%; Pred. No. 34;
 Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 QY 1 TCTAGTGTCTGTCGAGGCAT 19
 |||||
 Db 29 TCTAGTGTCTGTCGAGGAAT 11
 RESULT 7
 AAA12025/c
 ID AAA12025 standard; DNA; 72 BP.
 XX AC AAA12025;
 XX 14-AUG-2000 (first entry)
 XX Murine PGK HRE-containing enhancer element MT PGK18+++ DNA.
 DE HRE; hypoxia response element; hypoxia-inducible factor; HIF; vasotropic;
 KW cardiant; cytostatic; antiarthritic; gene therapy; ischaemia; arthritis;
 KW cardiovascular disease; peripheral arterial disease; cancer;
 KW phosphoglycerate kinase; PGK; murine; enhancer; ds.
 XX Mus sp.
 OS WO200017371-A1.
 XX 30-MAR-2000.
 XX 22-SEP-1999; 99WO-GB03181.
 XX 23-SEP-1998; 98WO-GB02885.
 XX 28-JAN-1999; 99GB-0001906.
 XX 16-FEB-1999; 99GB-0003538.
 XX (OXFO-) OXFORD BIOMEDICA UK LTD.
 PA Binley KM, Naylor S;
 PI

DR WPI; 2000-283595/24.
XX
PT Novel polynucleotide constructs comprising at least two repeats of a
PT hypoxia response element useful for driving expression of nucleic acids
PT of interest in a cell under hypoxic conditions
PS Example 6; Page 79; 155pp; English.
XX
CC This invention describes novel polynucleotide comprising at least 2
CC repeats of a hypoxia response element (HRE), where the hypoxia-inducible
CC factor (HIF) consensus binding sites within each of the 2 repeats are
CC separated by a spacer of at least 20 contiguous nucleotides. The products
CC of the invention have vasotropic, cardiac, cytostatic and antiarthritic
CC activity and can be used for gene therapy. The polynucleotides are useful
CC for delivering nucleic acids of interest to mammalian cells. Lentiviral
CC vectors are responsive to hypoxic agents and to agents that mimic
CC hypoxia. This regulation can be harnessed in vitro to enhance the
CC production of the vector and can be used in vivo to regulate the
CC expression in response to a physiological signal. The vectors have
CC utility in disease, where ischemia, including hypoxia, is a feature,
CC e.g. cardiovascular disease, peripheral arterial disease, cancer and
CC arthritis. The novel regulatory construct is capable of driving very high
CC levels of transcription under conditions of hypoxia whilst providing only
CC low basal levels of transcription under normal oxygen conditions. The
CC polynucleotide construct targets cells within a tumor mass that are under
CC conditions of hypoxia without affecting normal surrounding tissue. This
CC sequence represents a murine phosphoglycerate kinase (PGK) HRE derived
CC enhancer which is described in the method of the invention.
XX
SQ Sequence 72 BP; 16 A; 25 C; 12 G; 19 T; 0 other;
Query Match 72.5%; Score 17.4; DB 21; Length 72;
Best Local Similarity 94.7%; Pred. No. 34;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 TCTAGTGTCTGTCGAGGCAT 19
| | | | | | | | | | | | | | | | | | | | | |
Db 29 TCTAGTGTCTGTCGAGGCAT 11
RESULT 8
ID AAN70749/C
XX AAN70749 standard; DNA; 39 BP.
AC
XX AAN70749;
DT 03-OCT-2002 (updated)
DT 04-FEB-1991 (first entry)
XX
DE Sequence of portion of Drosophila heat-shock hybrid gene on p17-lys
DE showing heat shock protein (Hsp70) promoter.
XX
KW Gene expression; transcription control element;
KW Drosophila 70kD heat shock protein promoter; ss;
KW chicken lysozyme gene; human hsp70 gene.
OS
XX Drosophila melanogaster.
XX
FH Key Location/Qualifiers
FT misc_feature 1..14
FT /tag= a
FT /note="Hsp70 gene sequence"
XX
PN W08705935-A.
XX
XX 08-OCT-1987.
PD
XX
PF 03-APR-1987; 87WO-US00805.
XX
PR 04-APR-1986; 86US-0848657.
XX
XX (BATT) BATTELLE MEMORIAL INST.
PA
XX

PI Bromley P, Voellmy R;
XX
DR WPI; 1987-291648/41.
XX
PT DNA contg. structural gene and heat shock promoter - esp. from
PT Drosophila, allowing induction of protein expression in prokaryotic
PT or eukaryotic cells.
XX
PS Example; Fig 13; 105pp; English.
XX
CC The sequence in AAN70743 was cut from plasmid p51 and inserted into
CC pMC 1403 in front of its incomplete beta-galactosidase gene to form
CC recombinant plasmid pRV15 (specifically claimed) which was able to
CC express beta-galactosidase. pRV15 was digested with SmaI and SalI to
CC give a 7 kbp segment and this ligated with pSVod to form plasmid
CC 520. Separately, plasmid pR81 was digested with HindIII and NcoI.
CC The resulting fragments were ligated with a PvuII-NcoI fragment of
CC p520 to form the specifically claimed plasmid pR84 (AAN70744). The
CC heat shock promoters are used for the inducible expression of genes
CC encoding secretable and non-secretable proteins. A 1.5 kb fragment
CC from p522-lys and ligated to a 1.2 kb long fragment contg. a
CC functional heat shock promoter from p17 to give p17-lys.
CC (Updated on 03-OCT-2002 to add missing OS field.)
XX
SQ Sequence 39 BP; 12 A; 11 C; 10 G; 6 T; 0 other;
Query Match 69.2%; Score 16.6; DB 8; Length 39;
Best Local Similarity 82.6%; Pred. No. 78;
Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
QY 2 CTAGTGTCTGTCGAGGCATCTAGT 24
| | | | | | | | | | | | | | | | | | | | | |
Db 26 CCAGTGTCTGTCAGAGCTCTGTGT 4
RESULT 9
ID AAL25330/C
XX AAL25330 standard; cDNA; 111 BP.
AC AAL25330;
XX
DT 07-DEC-2001 (first entry)
XX
DE Human breast cancer expressed polynucleotide 17787.
XX
DE Human; breast cancer; cell marker; cytostatic; ss.
KW
XX Homo sapiens.
XX
XX W0200151628-A2.
XX
XX 19-JUL-2001.
XX
PF 10-JAN-2001; 2001WO-US00798.
XX
XX 14-JAN-2000; 2000US-0176077.
PR 14-MAR-2000; 2000US-0189167.
PR 24-MAR-2000; 2000US-0192099.
PR 29-MAR-2000; 2000US-0193480.
PR 15-MAY-2000; 2000US-0205230.
PR 09-JUN-2000; 2000US-0211315.
PR 25-JUL-2000; 2000US-0220534.
XX
XX (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
XX
XX Lillie J, Xu Y, Wang Y, Steinmann K;
XX
XX WPI; 2001-451856/48.
XX
PT New peptide useful as a marker for the diagnosis of breast cancer
XX
PS Claim 1; Page 3293; 3695pp; English.

XX The invention relates to human breast cancer expressed polynucleotides
 CC (AAL07544-AAL26789) and methods of assessing whether a patient is
 CC afflicted with breast cancer by examining the correlation between the
 CC expression of certain markers and the cancerous state of breast cells.
 CC The polynucleotides and encoded polypeptides are potential markers for
 CC detecting, diagnosing, monitoring, characterising treating and
 CC potentially preventing breast cancer. The polynucleotides and encoded
 CC polypeptides are also useful for isolating compounds with cytostatic
 CC activity.
 XX
 SQ Sequence 111 BP; 23 A; 30 C; 39 G; 19 T; 0 other;

Query Match 69.2%; Score 16.6; DB 22; Length 111;
 Best Local Similarity 82.6%; Pred. No. 91;
 Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 2 CTAGTCTCGTCAGGCATCTAGT 24
 DB 105 CTAGTCTCGTCAGGCATCTCGT 83

RESULT 10
 AAL16487/c
 ID AAL16487 standard; cDNA; 115 BP.

XX AC AAL16487;

XX DT 07-DEC-2001 (first entry)

XX DE Human breast cancer expressed polynucleotide 8944.

XX KW Human; breast cancer; cell marker; cytostatic; ss.

XX OS Homo sapiens.

XX PN WO200151628-A2.

XX PD 19-JUL-2001.

XX PF 10-JAN-2001; 2001WO-US00798.

XX PR 14-JAN-2000; 2000US-0176077.

XX PR 14-MAR-2000; 2000US-0189167.

XX PR 24-MAR-2000; 2000US-0192099.

XX PR 29-MAR-2000; 2000US-0193480.

XX PR 15-MAY-2000; 2000US-0205230.

XX PR 09-JUN-2000; 2000US-0211315.

XX PR 25-JUL-2000; 2000US-0220534.

XX (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.

XX Lillie J, Xu Y, Wang Y, Steinmann K;

XX WPI; 2001-451856/48.

XX New peptide useful as a marker for the diagnosis of breast cancer

XX Claim 1; Page 1618; 3695pp; English.

XX The invention relates to human breast cancer expressed polynucleotides
 CC (AAL07544-AAL26789) and methods of assessing whether a patient is
 CC afflicted with breast cancer by examining the correlation between the
 CC expression of certain markers and the cancerous state of breast cells.
 CC The polynucleotides and encoded polypeptides are potential markers for
 CC detecting, diagnosing, monitoring, characterising treating and
 CC potentially preventing breast cancer. The polynucleotides and encoded
 CC polypeptides are also useful for isolating compounds with cytostatic
 CC activity.

XX SQ Sequence 115 BP; 24 A; 30 C; 41 G; 20 T; 0 other;

Query Match 69.2%; Score 16.6; DB 22; Length 115;

Best Local Similarity 82.6%; Pred. No. 91;
 Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 2 CTAGTCTCGTCAGGCATCTAGT 24
 DB 109 CTAGTCTCGTCAGGCATCTCGT 87

RESULT 11
 ABV14349/c
 ID ABV14349 standard; cDNA; 192 BP.

XX AC ABV14349;

XX DT 13-SEP-2002 (first entry)

XX DE Human prostate expression marker cDNA 14340.

XX KW Human; prostate cancer; cytostatic; carcinogen; pharmacodynamic marker;
 KW pharmacogenomic marker; gene; ss.

XX OS Homo sapiens.

XX PN WO200160860-A2.

XX PD 23-AUG-2001.

XX PF 20-FEB-2001; 2001WO-US05171.

XX PR 17-FEB-2000; 2000US-183319P.

XX PR 16-MAR-2000; 2000US-189862P.

XX PR 25-MAY-2000; 2000US-207454P.

XX PR 09-JUN-2000; 2000US-211314P.

XX PR 18-JUL-2000; 2000US-219007P.

XX PR 13-DEC-2000; 2000US-255281P.

XX (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.

XX Schlegel R, Endege WO, Monahan JE;

XX WPI; 2001-662795/76.

XX Novel isolated nucleic acid molecule associated with cancerous state of
 PT prostate cells and correlating with presence of prostate cancer, useful
 PT for detecting presence of prostate cancer, stage of prostate cancer

PS Claim 1; Page 2394; 11750pp; English.

XX The invention relates to an isolated nucleic acid molecule (I) comprising
 CC a nucleotide sequence given in Tables 1-9 (ABV00010-ABV62213) of the
 CC specification or its complement. (I) is useful for:

CC (a) assessing whether a patient is afflicted with prostate cancer;

CC (b) monitoring the progression of prostate cancer in a patient;

CC (c) assessing the efficacy of a test compound to inhibit prostate
 CC cancer in a patient;

CC (d) assessing the efficacy of a therapy for inhibiting prostate cancer
 CC in a patient;

CC (e) selecting a composition for inhibiting prostate cancer in a patient;

CC (f) assessing the prostate cell carcinogenic potential of a compound;

CC (g) determining whether prostate cancer has metastasized in a patient;

CC (h) assessing the aggressiveness or indolence of prostate cancer in a
 CC patient;

CC (i) is also useful as a pharmacodynamic or pharmacogenomic marker.

XX SQ Sequence 192 BP; 38 A; 50 C; 70 G; 34 T; 0 other;

Query Match 69.2%; Score 16.6; DB 23; Length 192;
 Best Local Similarity 82.6%; Pred. No. 98;
 Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 2 CTAGTCTCGTCAGGCATCTAGT 24
 DB 186 CTAGTCTCGTCAGGCATCTCGT 164

Human; prostate cancer; cytostatic; carcinogen; pharmacodynamic marker; pharmacogenomic marker; gene; ss.

Homo sapiens.

WO200160860-A2.

23-AUG-2001.

20-FEB-2001; 2001WO-US05171.

17-FEB-2000; 2000US-183319P.

16-MAR-2000; 2000US-189862P.

25-MAY-2000; 2000US-207454P.

09-JUN-2000; 2000US-211314P.

18-JUL-2000; 2000US-219007P.

13-DEC-2000; 2000US-255281P.

(MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.

Schlegel R, Endege WO, Monahan JE;

WPI; 2001-662795/76.

Novel isolated nucleic acid molecule associated with cancerous state of prostate cells and correlating with presence of prostate cancer, useful for detecting presence of prostate cancer, stage of prostate cancer -

Claim 1; Page 8790; 11750pp; English.

The invention relates to an isolated nucleic acid molecule (I) comprising a nucleotide sequence given in Tables 1-9 (ABV00010-ABV62213) of the specification or its complement. (I) is useful for:

(a) assessing whether a patient is afflicted with prostate cancer;

(b) monitoring the progression of prostate cancer in a patient;

(c) assessing the efficacy of a test compound to inhibit prostate cancer in a patient;

(d) assessing the efficacy of a therapy for inhibiting prostate cancer in a patient;

(e) selecting a composition for inhibiting prostate cancer in a patient;

(f) assessing the prostate cell carcinogenic potential of a compound;

(g) determining whether prostate cancer has metastasized in a patient;

(h) assessing the aggressiveness or indolence of prostate cancer in a patient;

(I) is also useful as a pharmacodynamic or pharmacogenomic marker.

Sequence 450 BP; 93 A; 118 C; 143 G; 96 T; 0 other;

Query Match 69.2%; Score 16.6; DB 23; Length 450;

Best Local Similarity 82.6%; Pred. No. 1.1e+02;

Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 2 CTAGTGTCTGTCAGGCATCTAGT 24

||||| ||||| ||||| ||

Db 228 CTAGTCTCTGTCACGGATCTCGT 206

RESULT 14

ABV35433/C

ID ABV35433 standard; cDNA; 479 BP.

XX

AC ABV35433;

XX

XX

16-SEP-2002 (first entry)

XX

DE Human prostate expression marker cDNA 35424.

XX

KW Human; prostate cancer; cytostatic; carcinogen; pharmacodynamic marker; pharmacogenomic marker; gene; ss.

XX

OS Homo sapiens.

XX

PN WO200160860-A2.

RESULT 12

AAC09061/C

ID AAC09061 standard; cDNA; 202 BP.

XX

AC AAC09061;

XX

DT 06-OCT-2000 (first entry)

XX

DE Human secreted protein 5' EST, SEQ ID NO: 13136.

XX

KW Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;

XX

KQ gene therapy; chromosome mapping; ss.

XX

OS Homo sapiens.

XX

PN EP1033401-A2.

XX

PD 06-SEP-2000.

XX

PF 21-FEB-2000; 2000EP-0200610.

XX

PR 26-FEB-1999; 99US-0122487.

XX

PA (GEST) GENSET.

XX

PI Dumas Milne Edwards J, Duclert A, Giordano J;

XX

DR WPI; 2000-500381/45.

XX

PT New nucleic acid that is a 5' expressed sequence tag (5' EST) for

PT obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for

PT diagnostic, forensic, gene therapy and chromosome mapping procedures -

XX

PS Claim 1; SEQ ID 13136; 71pp + CD-ROM; English.

XX

CC The present sequence is one of a large number of 5' ESTs derived from

CC mRNAs encoding secreted proteins. No ORF has yet been conclusively

CC identified within the present sequence. The 5' ESTs were prepared from

CC total human RNAs or polyA+ RNAs derived from 30 different tissues. EST

CC sequences usually correspond mainly to the 3' untranslated region (UTR)

CC of the mRNA because they are often obtained from oligo-dT primed cDNA

CC libraries. Such ESTs are not well suited for isolating cDNA sequences

CC derived from the 5' ends of mRNAs and even in those cases where longer

CC cDNA sequences have been obtained, the full 5' UTR is rarely included.

CC 5' ESTs are derived from mRNAs with intact 5' ends and can therefore be

CC used to obtain full length cDNAs and genomic DNAs. 5' ESTs are also used

CC in diagnostic, forensic, gene therapy and chromosome mapping procedures.

CC They are used to obtain upstream regulatory sequences and to design

CC expression and secretion vectors.

XX

SQ Sequence 202 BP; 40 A; 53 C; 74 G; 35 T; 0 other;

Query Match 69.2%; Score 16.6; DB 21; Length 202;

Best Local Similarity 82.6%; Pred. No. 99;

Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 2 CTAGTGTCTGTCAGGCATCTAGT 24

||||| ||||| ||||| ||

Db 164 CTAGTCTCTGTCACGGATCTCGT 142

RESULT 13

ABV44262/C

ID ABV44262 standard; cDNA; 450 BP.

XX

AC ABV44262;

XX

XX

16-SEP-2002 (first entry)

XX

DE Human prostate expression marker cDNA 44253.

XX

```
XX PD 23-AUG-2001.
XX PR
XX PF
XX PI
XX XX 20-FEB-2001; 2001WO-US05171.
XX PR
XX PR 17-FEB-2000; 2000US-183319P.
XX PR 16-MAR-2000; 2000US-189862P.
XX PR 25-MAY-2000; 2000US-207454P.
XX PR 09-JUN-2000; 2000US-211314P.
XX PR 18-JUL-2000; 2000US-219007P.
XX PR 13-DEC-2000; 2000US-255281P.
XX PA (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
XX PI Schlegel R, Endege WO, Monahan JE;
XX DR WPI; 2001-662795/76.
XX PT Novel isolated nucleic acid molecule associated with cancerous state of
XX PT prostate cells and correlating with presence of prostate cancer, useful
XX PT for detecting presence of prostate cancer, stage of prostate cancer
XX PS Claim 1; Page 7380; 11750pp; English.
XX CC The invention relates to an isolated nucleic acid molecule (I) comprising
XX CC a nucleotide sequence given in Tables 1-9 (ABV00010-ABV62213) of the
XX CC specification or its complement. (I) is useful for:
XX CC (a) assessing whether a patient is afflicted with prostate cancer;
XX CC (b) monitoring the progression of prostate cancer in a patient;
XX CC (c) assessing the efficacy of a test compound to inhibit prostate
XX CC cancer in a patient;
XX CC (d) assessing the efficacy of a therapy for inhibiting prostate cancer
XX CC in a patient;
XX CC (e) selecting a composition for inhibiting prostate cancer in a patient;
XX CC (f) assessing the prostate cell carcinogenic potential of a compound;
XX CC (g) determining whether prostate cancer has metastasized in a patient;
XX CC (h) assessing the aggressiveness or indolence of prostate cancer in a
XX CC patient;
XX CC (I) is also useful as a pharmacodynamic or pharmacogenomic marker.
XX SQ Sequence 479 BP; 99 A; 122 C; 155 G; 103 T; 0 other;

Query Match 69.28; Score 16.6; DB 23; Length 479;
Best Local Similarity 82.6%; Pred. No. 1.1e+02;
Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 2 CTAGTCTCGTCGAGGCATCTAGT 24
Db 227 CTAGTCTCGTCGAGGCATCTCGT 205

RESULT 15
AAS10807
ID AAS10807 standard; cDNA; 726 BP.
XX AC AAS10807;
XX XX
XX DT 24-OCT-2001 (first entry)
XX DE Human Janus kinase 3 (JAK3) partial cDNA sequence, 3-2 SK.
XX KW Human; Janus kinase 3; JAK3; ss; JAK/STAT inhibitor; 3-2 SK;
XX KW signal transducer and activator of transcription; osteoarthritis;
XX KW degenerative joint disease; rheumatoid arthritis; leprosy; asthma;
XX KW cancer; tumour; leukaemia.
XX OS Homo sapiens.
XX PN WO200152892-A2.
XX PD 26-JUL-2001.
XX PF 22-JAN-2001; 2001WO-US02033.
```

```
XX PR 24-JAN-2000; 2000US-0177872.
XX PR 28-NOV-2000; 2000US-0723490.
XX PA (GENZ ) GENZYME CORP.
XX XX Vasios G;
XX DR WPI; 2001-465338/50.
XX PT Use of inhibitors of Janus kinase/signal transducers and activators of
XX PT transcription for inhibiting onset and progression of degenerative
XX PT joint diseases or disorders such as osteoarthritis, rheumatoid
XX PT arthritis
XX PS Example 4; Fig 3; 55pp; English.
XX CC The sequence is a partial sequence from a chondrocyte cDNA encoding
XX CC human Janus Kinase 3 (JAK3). The invention relates to the use of
XX CC JAK/STAT (Janus Kinase/signal transducer and activator of
XX CC transcription) inhibitors other than debromohymenialdisine (DBH) and
XX CC hymenialdisine (H) for inhibiting the progression or the likelihood of
XX CC developing diseases involving cartilage degradation, and for regulating
XX CC the expression of pro-inflammatory agents or cytokines in a chondrocyte,
XX CC and cartilage degrading enzymes in a cell. A JAK3/STAT inhibitor of the
XX CC invention is useful for inhibiting progression or likelihood of
XX CC developing osteoarthritis or rheumatoid arthritis. The inhibitor is also
XX CC useful for treating other JAK/STAT-mediated diseases or disorders,
XX CC including T cell-mediated disorders, mast cell-mediated disorders,
XX CC type 2 (cytokine hypersensitivity) disorders, B cell lymphoma, and
XX CC myeloid diseases. T cell-mediated disorders include human T cell
XX CC leukaemia/lymphoma virus (HTLV)-1, Sdzory's syndrome, c-abl
XX CC transformation, natural killer-like T cell lymphomas (NK-like tumours)
XX CC and graft-vs-host disease; cytokine hypersensitivity disorders include
XX CC Leishmaniasis, leprosy, allergy and viral infections; mast cell-mediated
XX CC disorders include hay fever, asthma, hives and anaphylaxis; and
XX CC leukaemias and lymphomas include acute lymphocytic and lymphoblastic
XX CC leukaemias, B cell lymphomas and leukaemias of myeloid origin. DBH and H
XX CC are useful as therapeutic agents in cancers in which JAK3 plays a role
XX CC in the initiation or progression of tumourigenesis.
XX SQ Sequence 726 BP; 151 A; 224 C; 204 G; 147 T; 0 other;

Query Match 69.28; Score 16.6; DB 22; Length 726;
Best Local Similarity 82.6%; Pred. No. 1.2e+02;
Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 2 CTAGTCTCGTCGAGGCATCTAGT 24
Db 677 CAAGTCTCGTACCGCATCTCGT 699

Search completed: January 3, 2003, 23:20:46
JOB time : 42.1262 secs
```

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 3, 2003, 22:54:17 : Search time 319.117 Seconds
(without alignments)
1218.024 Million cell updates/sec

Title: US-09-787-562-11

Perfect score: 24

Sequence: 1 tctagtctgtgcaggcatctagt 24

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- 1: em_estba:**
- 2: em_esthum:**
- 3: em_estin:**
- 4: em_estmu:**
- 5: em_estov:**
- 6: em_estpl:**
- 7: em_estro:**
- 8: em_htc:**
- 9: gb_estl:**
- 10: gb_est2:**
- 11: gb_htc:**
- 12: gb_est3:**
- 13: gb_est4:**
- 14: gb_est5:**
- 15: em_estfun:**
- 16: em_estom:**
- 17: gb_gss:**
- 18: em_gss_hum:**
- 19: em_gss_inv:**
- 20: em_gss_pln:**
- 21: em_gss_vrt:**
- 22: em_gss_fun:**
- 23: em_gss_man:**
- 24: em_gss_mus:**
- 25: em_gss_other:**
- 26: em_gss_pro:**
- 27: em_gss_rod:**

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	20.4	85.0	529	BU011601	BU011601 QGJ16K24.
2	20.4	85.0	682	14 BQ843759	BQ843759 QGJ11I20.
3	18.2	75.8	567	17 BH206532	BH206532 Sml-46N18
4	17.8	74.2	297	17 AQ177064	AQ177064 HS-2216.B
5	17.8	74.2	503	17 BH729239	BH729239 BOMFY07TF
6	17.6	73.3	206	12 BF086292	BF086292 QV3-GN006

7	17.6	73.3	207	14	BQ366039	BQ366039 QV3-GN006
8	17.6	73.3	244	9	AA309771	AA309771 EST180630
9	17.6	73.3	300	17	AZ827575	AZ827575 ZM0104005
10	17.6	73.3	520	12	BE930148	BE930148 QV3-GN006
11	17.6	73.3	582	9	AL705176	AL705176 DKF26868
12	17.6	73.3	607	17	AQ882182	AQ882182 HS-5358.B
13	17.6	73.3	750	12	BG830180	BG830180 602764757
14	17.6	73.3	812	9	AL521259	AL521259 AU521259
15	17.6	73.3	843	9	AU133085	AU133085 AU133085
16	17.6	73.3	875	9	AU125595	AU125595 AU125595
17	17.6	73.3	875	14	BQ215175	BQ215175 AGENCOURT
18	17.6	73.3	1016	13	BM543250	BM543250 AGENCOURT
19	17.6	73.3	1105	13	BM458582	BM458582 AGENCOURT
20	17.6	73.3	1194	12	BF780452	BF780452 602103972
21	17.2	71.7	109	10	AW994796	AW994796 RCI-BN003
22	17.2	71.7	283	13	BI007122	BI007122 QV3-RN006
23	17.2	71.7	389	17	AG019642	AG019642 Homo sapi
24	17.2	71.7	399	17	BH758471	BH758471 SALK 0199
25	17.2	71.7	417	14	BU024781	BU024781 QHFG606.Y
26	17.2	71.7	470	10	AW615392	AW615392 hg82d07.x
27	17.2	71.7	499	17	BH039133	BH039133 RPCI-24-2
28	17.2	71.7	544	17	AQ678125	AQ678125 HS-5528.B
29	17.2	71.7	561	17	AZ012741	AZ012741 RPCI-23-3
30	17.2	71.7	682	17	AQ768930	AQ768930 HS-3160.B
31	17.2	71.7	684	17	AG071473	AG071473 Pan trogl
32	17.2	71.7	712	12	BG431040	BG431040 602498653
33	17.2	71.7	832	17	AQ897762	AQ897762 HS-3153.A
34	17.2	71.7	879	14	BQ071181	BQ071181 AGENCOURT
35	17.2	71.7	891	12	BF672432	BF672432 ENTID73TF
36	17.2	71.7	1019	17	AZ688328	AZ688328 ENTID73TF
37	17.2	71.7	1113	13	BM043304	BM043304 603619567
38	17.2	71.7	1392	12	BF029841	BF029841 601556785
39	16.8	70.0	192	10	AW893929	AW893929 RCI-NN002
40	16.8	70.0	346	17	AZ774173	AZ774173 2M0003115
41	16.8	70.0	370	10	BE332565	BE332565 us24b04.x
42	16.8	70.0	433	10	BE335885	BE335885 us24b04.x
43	16.8	70.0	497	17	AQ298884	AQ298884 HS-3157.B
44	16.8	70.0	718	17	BH879219	BH879219 ht45a12.b
45	16.8	70.0	764	17	AQ870122	AQ870122 nbeb0036p

ALIGNMENTS

RESULT 1	BU011601	529 bp	mRNA	linear	EST 22-AUG-2002
LOCUS	QGJ16K24.yg.abl QG_EFGHJ	lettuce	serriola	Lactuca sativa	cDNA clone
DEFINITION	QGJ16K24, mRNA sequence.				
ACCESSION	BU011601				
VERSION	BU011601.1	GI:22445996			
KEYWORDS	EST				
SOURCE	Lactuca sativa.				
ORGANISM	Lactuca sativa				
REFERENCE	1 (bases 1 to 529)				
AUTHORS	Kozik, A., Michelmore, R.W., Knapp, S., Matvienko, M., Riesberg, L., Lin, H., van Damme, M., Lavelle, D., Chevalier, P., Ziegler, J., Ellison, P., Kolkman, J., Slabaugh, M.S., Livingston, K., Zhou, Y., Lai, Z., Church, S., Jackson, L. and Bradford, K.				
TITLE	Lettuce and Sunflower ESTs from the Compositae Genome Project				
JOURNAL	Unpublished (2002)				
COMMENT	Contact: Alexander Kozik [R.W.Michelmore] Department of Vegetable Crops, R.W.Michelmore Lab University of California at Davis (UCD) Asmadson Hall, UCD, Davis, CA 95616, USA Tel: 1-(530)-742-1742 Fax: 1-(530)-752-9659 Email: akozik@ucdavis.org [michelmore@vegmail.ucdavis.edu]				


```

Query Match      75.8%; Score 18.2; DB 17; Length 567;
Best Local Similarity 87.0%; Pred. No. 6.4e+02;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 TCTAGTGTCTGTCAGGCATCTAG 23
||||| ||||| ||||| ||||| |||||
Db 421 TCTAGTGTCTTGGAGTCATCTAG 399

RESULT 4
AQ177064/c
LOCUS          297 bp      DNA      linear      GSS 17-OCT-1998
DEFINITION    HS_2216_B2_E08_MF_CIT Approved Human Genomic Sperm Library D Homo
               sapiens genomic clone Plate=2216 Col=16 Row=J, DNA sequence.
ACCESSION     AQ177064
VERSION       AQ177064.1 GI:3574431
KEYWORDS      GSS.
SOURCE        human.
ORGANISM      Homo sapiens
               Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
               Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE     1 (bases 1 to 297)
AUTHORS       Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,
               Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and
               Hood,L.
TITLE         Sequence-tagged connectors: A sequence approach to mapping and
               scanning the human genome
JOURNAL       Proc. Natl. Acad. Sci. U.S.A. 96 (17), 9739-9744 (1999)
MEDLINE       99380589
COMMENT       Contact: Mahairas GG, Wallace JC, Hood L
               High Throughput Sequencing Center
               University of Washington
               401 Queen Anne Avenue North, Seattle, WA 98109, USA
               Tel: (206) 616-3618
               Fax: (206) 616-3887
               Email: jwallace@u.washington.edu
               Sequence Tagged Connector
               Plate: 2216 row: J column: 16
               Class: BAC ends
               High quality sequence stop: 297.
FEATURES      Location/Qualifiers
               source          1..297
               /organism="Homo sapiens"
               /db_xref="taxon:9606"
               /clone="Plate=2216 Col=16 Row=J"
               /clone_lib="CIT Approved Human Genomic Sperm Library D"
               /sex="male"
               /note="Organ: sperm; Vector: pBelobAC11; BAC Clones in
               E-Coli DH10B"
BASE COUNT    84 a 80 c 66 g 67 t
ORIGIN
Query Match      74.2%; Score 17.8; DB 17; Length 297;
Best Local Similarity 90.5%; Pred. No. 7.5e+02;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 TCTAGTGTCTGTCAGGCATCT 21
||||| ||||| ||||| ||||| |||||
Db 221 TCTAGGGTCGGCAGGCATCT 201

RESULT 5
BH729239
LOCUS          503 bp      DNA      linear      GSS 20-FEB-2002
DEFINITION    BOMFY07TF BO_2_3_KB Brassica oleracea genomic clone BOMFY07, DNA
               sequence.
ACCESSION     BH729239
VERSION       BH729239.1 GI:18834634
KEYWORDS      GSS.
SOURCE        Brassica oleracea.
ORGANISM      Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

```

```

Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids II; Brassicales; Brassicaceae; Brassica.
1 (bases 1 to 503)
Town,C.D., Van Aken,S., Utterback,T. and Fraser,C.M.
Whole genome shotgun sequencing of Brassica oleracea
Unpublished (2001)
Other_GSSs: BOMFY07TR
Contact: Chris Town
TIGR
9712 Medical Center Drive, Rockville, MD 20850, USA.
Tel: 301-838-3523
Fax: 301-838-0208
Email: cdtown@tigr.org
DNA is from a doubled haploid provided by Tom Osborn.
Seq primer: TF
Class: sheared ends.
FEATURES      Location/Qualifiers
               source          1..503
               /organism="Brassica oleracea"
               /strain="TO1000DH3"
               /db_xref="taxon:3712"
               /clone="BOMFY07"
               /clone_lib="BO_2_3_KB"
               /note="Vector: pHOS1; Site 1: BstXI; 2-3 kb sheared
               genomic DNA inserted into pHOS1 using BstXI linkers"
BASE COUNT    178 a 85 c 72 g 168 t
ORIGIN
Query Match      74.2%; Score 17.8; DB 17; Length 503;
Best Local Similarity 90.5%; Pred. No. 9.3e+02;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 TCTAGTGTCTGTCAGGCATCT 21
||||| ||||| ||||| ||||| |||||
Db 407 TCGAGTGTCTGTCATCATCT 427

RESULT 6
BF086292
LOCUS          206 bp      mRNA      linear      EST 19-OCT-2000
DEFINITION    QV3-GN0065-120900-323-g02 GN0065 Homo sapiens cDNA, mRNA sequence.
ACCESSION     BF086292
VERSION       BF086292.1 GI:10892002
KEYWORDS      EST.
SOURCE        human.
ORGANISM      Homo sapiens
               Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
               Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE     1 (bases 1 to 206)
AUTHORS       Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
               Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
               Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
               Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
               ,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
               Simpson,A.J.
TITLE         Shotgun sequencing of the human transcriptome with ORF expressed
               sequence tags
JOURNAL       Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
MEDLINE       20202663
COMMENT       Contact: Simpson A.J.G.
               Laboratory of Cancer Genetics
               Ludwig Institute for Cancer Research
               Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
               Brazil
               Tel: +55-11-2704922
               Fax: +55-11-2707001
               Email: asimpson@ludwig.org.br
               This sequence was derived from the FAPESP/LICR Human Cancer Genome
               project. This entry can be seen in the following URL
               (http://www.ludwig.org.br/scripts/gethtml2.pl?tl=st2-QV3-GN0065-120
               900-323-g02&t3=2000-09-12&t4=1)
               Seq primer: puc 18 forward
               High quality sequence start: 6

```

High quality sequence stop: 206.

FEATURES

source

Location/Qualifiers
1. .206
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="GN0065"
/dev_stage="Adult"

/note="Organ: Placenta_normal; Vector: puc18; Site_1: SmaI
; Site_2: SmaI; A mini-library was made by cloning
products derived from ORESTES PCR (U.S. Letters Patent
application No. 196,716 - Ludwig Institute for Cancer
Research) profiles into the pUC 18 vector. Reverse
transcription of tissue mRNA and cDNA amplification were
performed under low stringency conditions."

BASE COUNT

ORIGIN

63 a 37 c 51 g 55 t

Query Match 73.3%; Score 17.6; DB 12; Length 206;

Best Local Similarity 83.3%; Pred. No. 8e+02;

Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 TCTAGTGTCTGTCAGGCACTAGT 24

||||| ||||| ||||| |||||

Db 94 TCTGGAGTCGTGGAGCAACTAGT 117

RESULT 7

BQ366039

LOCUS

DEFINITION

QV3-GN0061-270900-323-g02 GN0061 Homo sapiens cDNA, mRNA sequence.

ACCESSION

BQ366039.1 GI:21041551

VERSION

BQ366039.1

KEYWORDS

EST.

SOURCE

human.

ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.

1 (bases 1 to 207)

Dias Neto, E., Garcia Correa, R., Verjovski-Almeida, S., Briones, M.R.,

Nagai, M.A., da Silva, W. Jr., Zago, M.A., Bordin, S., Costa, F.F.,

Goldman, G.H., Carvalho, A.F., Matsukuma, A., Bata, G.S., Simpson, D.H.,

Brunstein, A., de Oliveira, P.S., Bucher, P., Jongeneel, C.V., O'Hare

, M.J., Soares, F., Brentani, R.R., Reis, L.F., de Souza, S.J. and

Simpson, A.J.

Shotgun sequencing of the human transcriptome with ORF expressed

sequence tags

Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)

20202663

Contact: Simpson A.J.G.

Laboratory of Cancer Genetics

Ludwig Institute for Cancer Research

Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,

Brazil

Tel: +55-11-2704922

Fax: +55-11-2707001

Email: asimpson@ludwig.org.br

This sequence was derived from the FAPESP/LICR Human Cancer Genome

project. This entry can be seen in the following URL

(http://www.ludwig.org.br/scripts/gethtml2.pl?tl=QV3&t2=QV3-GN0061-

270900-323-g02&t3=2000-09-27&t4=1)

Seq primer: puc 18 forward

High quality sequence start: 6

High quality sequence stop: 207.

Location/Qualifiers

1. .207

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone_lib="GN0061"

/dev_stage="Adult"

/note="Organ: placenta_normal; Vector: puc18; Site_1: SmaI

; Site_2: SmaI; A mini-library was made by cloning

products derived from ORESTES PCR (U.S. Letters Patent

application No. 196,716 - Ludwig Institute for Cancer

BASE COUNT

ORIGIN

64 a 37 c 51 g 55 t

Query Match

Best Local Similarity 83.3%; Score 17.6; DB 14; Length 207;

Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 TCTAGTGTCTGTCAGGCACTAGT 24

||||| ||||| ||||| |||||

Db 95 TCTGGAGTCGTGGAGCAACTAGT 118

RESULT 8

AA309771

LOCUS

DEFINITION

EST180630 Jurkat T-cells V Homo sapiens cDNA 5' end similar to

hypothetical protein KIAA0225, mRNA sequence.

AA309771

ACCESSION

AA309771.1 GI:1962100

VERSION

EST.

KEYWORDS

SOURCE

human.

ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.

1 (bases 1 to 244)

Adams, M.D., Kerlavage, A.R., Fleischmann, R.D., Fuldner, R.A., Bult

, C.J., Lee, N.H., Kirkness, E.F., Weinstock, K.G., Gocayne, J.D., White

, O., Sutton, G., Blake, J.A., Brandon, R.C., Man-Wai, C., Clayton, R.A.,

Cline, T.R., Cotton, M.D., Earle-Hughes, J., Fine, L.D., Fitzgerald

, L.M., Fitzhugh, W.M., Fritchman, J.L., Geoghagen, N.S., Glodek, A.,

Gnehm, C.L., Hanna, M.C., Hedblom, E., Hinkle, P.S. Jr., Kelley, J.M.,

Kelley, J.C., Liu, L.-I., Marmaros, S.M., Merrick, J.M.,

Moreno-Palancas, R.F., McDonald, L.A., Nguyen, D.T., Pelligrino, S.M.,

Phillips, C.A., Ryder, S.E., Scott, J.L., Saudek, D.M., Shirley, R.,

Small, K.V., Springs, F.A., Utterback, T.A., Weidman, J.F., Li, Y.,

Bednarek, D.P., Cao, L., Cepeda, M.A., Coleman, T.A., Collins, E.J.,

Dimke, D., Feng, D.-F., Ferrie, A., Fischer, C., Hastings, G.A., He, W.W.,

Hu, J.S., Greene, J.M., Gruber, J., Hudson, P., Kim, A.K., Kozak, D.L.,

Kunsch, C., Greene, J.M., Li, H., Weissner, P.S., Olsen, H., Raymond, L.,

Wei, Y.F., Wang, J., Xu, C., Yu, G.L., Ruben, S.M., Dillion, P.J., Fannon

, M.R., Rosen, C.A., Haseitine, W.A., Fields, C., Fraser, C.M. and

Venter, J.C.

Initial assessment of human gene diversity and expression patterns

based upon 83 million nucleotides of cDNA sequence

Nature 377 (6547 Suppl), 3-174 (1995)

96026280

Other ESTs: THCL89719

Contact: Kerlavage, AR

Bioinformatics

The Institute for Genomic Research

9712 Medical Center Drive, Rockville, MD 20850 USA

Tel: 3018699056

Fax: 3018699423

Email: arkerlav@tigr.org

For clone availability, additional sequence and expression

information related to this EST, please check the TIGR Human Gene

Index (http://www.tigr.org/tdb/hgi/hgi.html)

Seq primer: M3 Reverse.

Location/Qualifiers

1. .244

/organism="Homo sapiens"

/db_xref="ATCC (inhost):155863"

/db_xref="taxon:9606"

/clone_lib="Jurkat T-cells V"

/cell_type="T-lymphocyte"

/note="Vector: pBluescript SK-; Site_1: EcoRI; Site_2:

XhoI"

73 a 46 c 59 g 63 t 3 others

BASE COUNT

ORIGIN

BASE COUNT
ORIGIN

```

LOCUS      AL705176      582 bp      mRNA      linear      EST 22-MAR-2002
DEFINITION DKFZp686N1034_r1 686 (synonym: hlcc3) Homo sapiens cDNA clone
ACCESSION  AL705176
VERSION     AL705176.1 GI:19688531
KEYWORDS   EST.
SOURCE      human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE  1 (bases 1 to 582)
AUTHORS   Ottenwaelder,B., Obermaier,B., Mewes,H.W., Mewes,H.W., Weil,B. and
            Wiemann,S.
TITLE      EST (Ottenwaelder,B., Obermaier,B., Mewes,H.W., Weil,B. and Wiemann
            S.)
JOURNAL    Unpublished (2001)
COMMENT    Contact: Ottenwaelder B
            MIPS
            Am Klopferstritz 18a D-82152 Martinsried, Germany
            This is the 5' sequence of the clone insert
            Clone from S. Wiemann, Molecular Genome Analysis, German Cancer
            Research Center (DKFZ); Email s.wiemann@dkfz-heidelberg.de;
            sequenced by Medigenomix (Martinsried/Germany) within the cDNA
            sequencing consortium of the German Genome Project. No sl sequence
            available.
            This clone (DKFZp686N1034) is available at the RZPD in Berlin.
            Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059
            Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de.
FEATURES   Location/Qualifiers
            1..582
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
            /clone="DKFZp686N1034"
            /clone_lib="686 (synonym: hlcc3)"
            /tissue_type="human skeletal muscle"
            /dev_stage="adult"
            /lab_host="DH10B"
            /note="Vector: pTriplex2; Site_1: SfiIA; Site_2: SfiIB;
            cDNA-collection"
BASE COUNT 168 a 123 c 137 g 154 t
ORIGIN
Query Match 73.3%; Score 17.6; DB 9; Length 582;
Best Local Similarity 83.3%; Pred. No. 1.2e+03;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 TCTAGTCTGTCGAGGCATCTAGT 24
    ||| ||||| ||||| |||||
Db 491 TCTGGAGTCTGTCGAGGCAACTAGT 514

RESULT 12
LOCUS      AQ882182      607 bp      DNA      linear      GSS 09-NOV-1999
DEFINITION HS_5358_B2_B06_T7A RPCI-11 Human Male BAC Library Homo sapiens
            genomic clone Plate=9126 Col=12 Row=8, DNA sequence.
ACCESSION  AQ882182
VERSION     AQ882182.1 GI:6313649
KEYWORDS   GSS.
SOURCE      human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE  1 (bases 1 to 607)
AUTHORS   Mahairas,G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,
            Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and
            Hood,L.
TITLE      Sequence-tagged connectors: A sequence approach to mapping and
            scanning the human genome
JOURNAL    Proc. Natl. Acad. Sci. U.S.A. 96 (17), 9739-9744 (1999)
MEDLINE    99380589
COMMENT    Contact: Mahairas GG, Wallace JC, Hood L
            High Throughput Sequencing Center

```

```

University of Washington
401 Queen Anne Avenue North, Seattle, WA 98109, USA
Tel: (206) 616-3618
Fax: (206) 616-3887
Email: jwallace@u.washington.edu
Clones are derived from the human BAC library RPCI-11. For BAC
library availability, please contact Pieter de Jong
(pieter@dejong.med.buffalo.edu). Clones may be purchased from
BACPAC Resources (http://bacpac.med.buffalo.edu/ordering_bac.htm)
or from Resear h Genetics (info@resgen.com). BAC end Web Server:
http://www.htsc.washington.edu
Plate: 9126 row: D column: 12
Seq primer: T7
Class: BAC ends
High quality sequence stop: 607.
FEATURES   Location/Qualifiers
            1..607
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
            /clone="Plate=9126 Col=12 Row=D"
            /clone_lib="RPCI-11 Human Male BAC Library"
            /sex="male"
            /note="Vector: pBACe3.6; Site_1: EcoRI; Site_2: EcoRI;
            Male blood DNA was isolated from one randomly chosen donor
            and partially digested with a combination of EcoRI and
            EcoRI Methylase. Size selected DNA was cloned into the
            pBACe3.6 vector at EcoRI sites"
BASE COUNT 170 a 116 c 109 g 204 t 8 others
ORIGIN
Query Match 73.3%; Score 17.6; DB 17; Length 607;
Best Local Similarity 83.3%; Pred. No. 1.2e+03;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 TCTAGTCTGTCGAGGCATCTAGT 24
    ||||| ||||| ||| |||||
Db 231 TCTAGTCTGTCGATGATCTAGT 254

RESULT 13
LOCUS      BG830180      750 bp      mRNA      linear      EST 22-MAY-2001
DEFINITION 602764757F1 NIH_MGC_42 Homo sapiens cDNA clone IMAGE:4906671 5',
            mRNA sequence.
ACCESSION  BG830180
VERSION     BG830180.1 GI:14177767
KEYWORDS   EST.
SOURCE      human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE  1 (bases 1 to 750)
AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
            National Institutes of Health, Mammalian Gene Collection (MGC)
            Unpublished (1999)
JOURNAL
COMMENT    Contact: Robert Strausberg, Ph.D.
            Email: cgabs-r@mail.nih.gov
            Tissue Procurement: ATCC
            cDNA Library Preparation: Ling Hong/Rubin Laboratory
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLCMI810 row: b column: 16
            High quality sequence stop: 723.
FEATURES   Location/Qualifiers
            1..750
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
            /clone="IMAGE:4906671"
            /clone_lib="NIH_MGC_42"
            /tissue_type="epithelioid carcinoma cell line"

```


/lab_host="DH10B (phage-resistant)"
 /note="Organ: pancreas; Vector: pOTB7; Site.1: XhoI;
 Site.2: EcoRI; cDNA made by oligo-dT priming.
 Directionally cloned into EcoRI/XhoI sites using the
 following 5' adaptor: GGCACGAG(G). Size-selected >500bp
 for average insert size 1.8Kb. Library constructed by Ling
 Hong in the laboratory of Gerald M. Rubin (University of
 California, Berkeley) using ZAP-cDNA synthesis kit
 (Stratagene) and Superscript II RT (Life Technologies).
 Note: this is a NIH_MGC Library. |"
 BASE COUNT 204 a 159 c 182 g 205 t
 ORIGIN

Query Match 73.3%; Score 17.6; DB 12; Length 750;
 Best Local Similarity 83.3%; Pred. No. 1.3e+03;
 Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
 QY 1 TCTAGTGTCTGTCGAGGCATCTAGT 24
 ||| | ||||| ||||| ||||| |||||
 Db 279 TCTGGAGTCGTGGAGGCAACTAGT 302

RESULT 14
 AL521259 812 bp mRNA linear EST 13-FEB-2001
 LOCUS AL521259 LTI_NFL004_NBC2 Homo sapiens cDNA clone CS0DB001YG12 5
 DEFINITION prime, mRNA sequence.
 ACCESSION AL521259
 VERSION AL521259.1 GI:12784752
 KEYWORDS EST.
 SOURCE human.

ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE 1 (bases 1 to 812)
 AUTHORS Li, W.B., Gruber, C., Jessee, J. and Polayes, D.
 TITLE Full-length cDNA libraries and normalization
 JOURNAL Unpublished (2001)
 COMMENT Contact: Genoscope
 Genoscope - Centre National de Sequencage
 BP 191 91006 EVRY cedex - France
 Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr.

FEATURES
 Source
 1..812
 Location/Qualifiers
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone_lib="CS0DB001YG12"
 /sex="male"
 /tissue_type="neuroblastoma cells"
 /lab_host="DH10B"

/note="Organ: brain; Vector: pCMVSPORT 6; 1st strand cDNA
 was primed with a NotI-oligo(dT) primer. Five prime end
 enriched, double-stranded cDNA was digested with Not I and
 cloned into the Not I and Eco RV sites of the pCMVSPORT 6
 vector. Library was normalized. Library was constructed
 by Life Technologies. Contact : Feng Liang Life
 Technologies, a division of Invitrogen 9800 Medical Center
 Drive Rockville, Maryland 20850, USA Fax : (1) 301 610
 8371 Email : fliang@lifetech.com URL :
 http://fulllength.invitrogen.com"
 BASE COUNT 220 a 174 c 195 g 221 t 2 others
 ORIGIN

Query Match 73.3%; Score 17.6; DB 9; Length 812;
 Best Local Similarity 83.3%; Pred. No. 1.4e+03;
 Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
 QY 1 TCTAGTGTCTGTCGAGGCATCTAGT 24
 ||| | ||||| ||||| ||||| |||||
 Db 72 TCTGGAGTCGTGGAGGCAACTAGT 95

RESULT 15
 AUI133085
 LOCUS AUI133085 NT2RP4 Homo sapiens cDNA clone NT2RP4001249 5', mRNA
 DEFINITION sequence.
 ACCESSION AUI133085
 VERSION AUI133085.1 GI:10993624
 KEYWORDS EST.
 SOURCE human.

ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

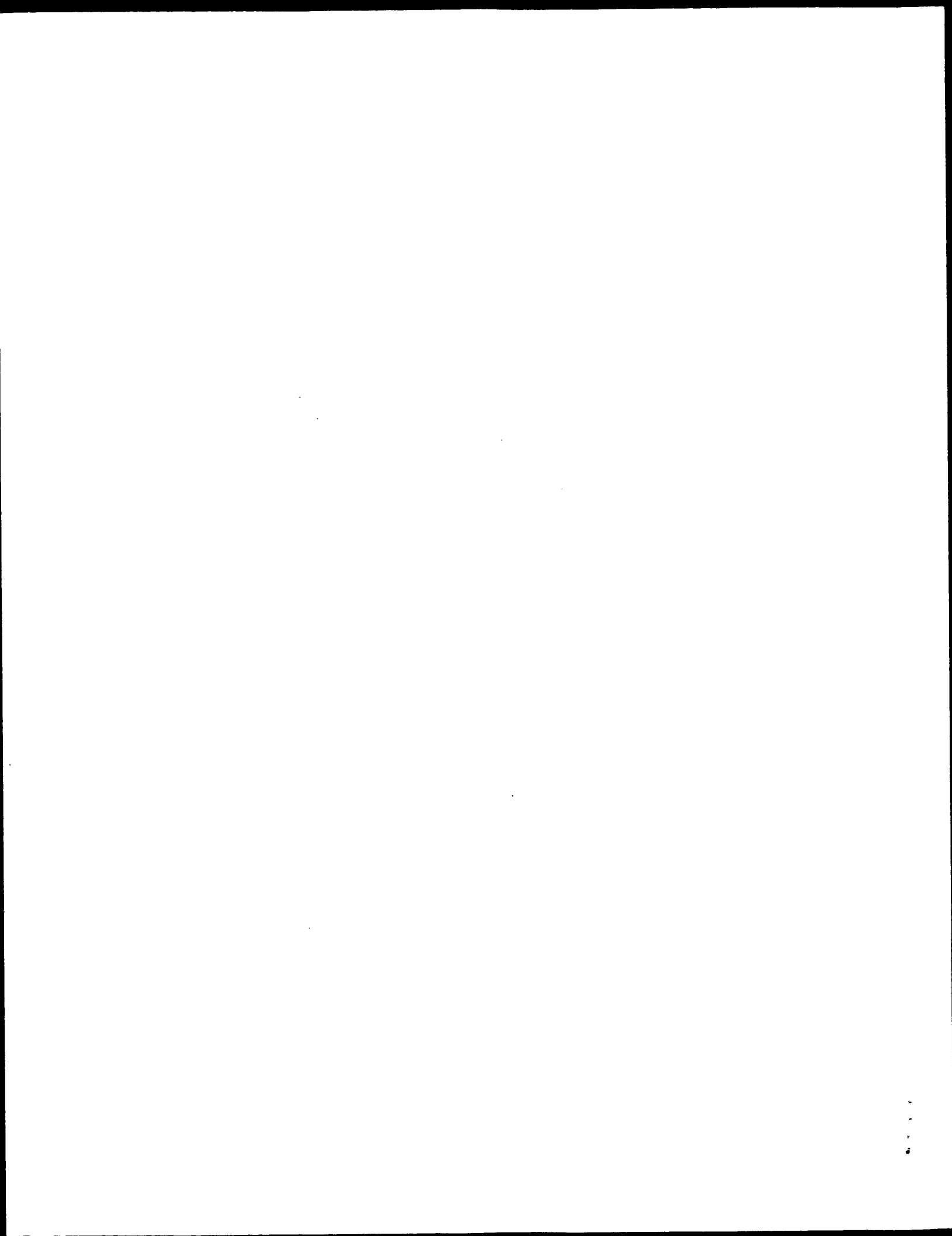
REFERENCE 1 (bases 1 to 843)
 AUTHORS Ota, T., Sugiyama, T., Ishii, S., Suzuki, Y., Saito, K., Yamamoto, J.,
 Nishikawa, T., Nakamura, Y., Nagai, T., Sugano, S., Masuho, Y. and
 Isogai, T.

TITLE HRI human cDNA project (Ota, T., Sugiyama, T., Ishii, S., Suzuki, Y.,
 Saito, K., Yamamoto, J., Nishikawa, T., Nakamura, Y., Nagai, T., Sugano
 S., Masuho, Y., Isogai, T.)
 JOURNAL Unpublished (2000)
 COMMENT Contact: Takao Isogai
 Genomics Laboratory
 Helix Research Institute
 1532-3 Yana, Kisarazu, Chiba 292-0812, Japan
 Tel: 81-438-52-3975
 Fax: 81-438-52-3986
 Email: genomics@hri.co.jp

HRI human cDNA project; 5' - & 3'-end one pass sequencing: Helix
 Research Institute; cDNA library construction: Department of
 Virology, Institute of Medical Science, University of Tokyo, and
 Helix Research Institute.
 FEATURES
 Location/Qualifiers
 1..843
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone_lib="NT2RP4001249"
 /cell_type="teratocarcinoma"
 /note="Vector: pME18SFL3; mRNA from NT2 neuronal precursor
 cells after 2-weeks retinoic acid (RA) induction"
 BASE COUNT 238 a 179 c 203 g 218 t 5 others
 ORIGIN

Query Match 73.3%; Score 17.6; DB 9; Length 843;
 Best Local Similarity 83.3%; Pred. No. 1.4e+03;
 Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
 QY 1 TCTAGTGTCTGTCGAGGCATCTAGT 24
 ||| | ||||| ||||| ||||| |||||
 Db 188 TCTGGAGTCGTGGAGGCAACTAGT 211

Search completed: January 4, 2003, 01:04:29
 Job time : 325.117 secs



GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 3, 2003, 22:53:16 ; Search time 8.10095 Seconds
(without alignments)
908.566 Million cell updates/sec

Title: US-09-787-562-11

Perfect score: 24

Sequence: 1 tctagtctgtcggcattagct 24

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 441362 seqs, 153338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- Issued Patents.NA.*
1: /cgn2_6/ptodata/2/ina/5A_COMB.seq.*
2: /cgn2_6/ptodata/2/ina/5B_COMB.seq.*
3: /cgn2_6/ptodata/2/ina/6A_COMB.seq.*
4: /cgn2_6/ptodata/2/ina/6B_COMB.seq.*
5: /cgn2_6/ptodata/2/ina/PTCUS_COMB.seq.*
6: /cgn2_6/ptodata/2/ina/backfiles1.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	16.6	69.2	3807	1	US-08-357-598-1
2	16.6	69.2	3807	2	US-09-003-289-1
3	16.6	69.2	3807	5	PCT-US95-16435-1
4	16.6	69.2	7559	2	US-08-250-848-2
C 5	16.6	66.7	17656	4	US-09-433-579-3
C 6	15.4	64.2	2057	3	US-09-008-303-1
7	15.4	64.2	246240	2	US-08-724-394A-20
8	15.4	64.2	246240	2	US-08-724-394A-21
9	15.4	64.2	246240	2	US-08-724-394A-22
C 10	15.2	63.3	2219	1	US-08-290-379A-7
11	15	62.5	594	4	US-09-328-111-155
C 12	15	62.5	1128	1	US-08-448-744-6
13	15	62.5	17056	4	US-09-245-041-3
C 14	14.8	61.7	822	1	US-07-644-372-1
15	14.8	61.7	3448	1	US-08-296-014A-3
16	14.8	61.7	3448	2	US-08-596-405-3
17	14.8	61.7	3448	2	US-08-877-620-3
18	14.8	61.7	4173	4	US-08-981-729-9
19	14.8	61.7	4173	4	US-08-981-446B-2
20	14.8	61.7	4182	1	US-08-296-014A-1
21	14.8	61.7	4182	2	US-08-596-405-1
22	14.8	61.7	4182	2	US-08-877-620-1
23	14.6	60.8	302	3	US-08-746-111-36
24	14.6	60.8	1251	4	US-09-072-596-319
25	14.6	60.8	3097	2	US-08-599-455B-1
26	14.6	60.8	3097	4	US-09-069-781B-1
27	14.6	60.8	3097	4	US-09-137-132-1

28	14.6	60.8	3097	4	US-08-864-564A-1	Sequence 1, Appli
29	14.6	60.8	3097	4	US-09-094-410-1	Sequence 1, Appli
30	14.6	60.8	3854	2	US-08-599-455B-42	Sequence 42, Appl
31	14.6	60.8	3854	4	US-09-069-781B-42	Sequence 42, Appl
32	14.6	60.8	3854	4	US-09-137-132-42	Sequence 42, Appl
33	14.6	60.8	3854	4	US-08-864-564A-42	Sequence 42, Appl
34	14.6	60.8	3854	4	US-09-094-410-42	Sequence 42, Appl
35	14.6	60.8	4264	2	US-08-784-649A-1	Sequence 1, Appli
36	14.6	60.8	4264	2	US-08-784-649A-5	Sequence 5, Appli
37	14.6	60.8	4646	1	US-08-181-471-2	Sequence 2, Appli
38	14.6	60.8	4669	2	US-08-583-276-18	Sequence 18, Appl
39	14.6	60.8	4669	2	US-08-752-447-1	Sequence 1, Appli
40	14.6	60.8	4669	4	US-09-316-167-1	Sequence 1, Appli
41	14.6	60.8	4669	6	5206352-3	Patent No. 5206352
42	14.6	60.8	4773	3	US-08-884-324-9	Sequence 9, Appli
43	14.6	60.8	6505	2	US-08-793-610-5	Sequence 5, Appli
44	14.6	60.8	6737	4	US-09-453-702B-76	Sequence 76, Appl
45	14.6	60.8	9318	2	US-08-793-610-6	Sequence 6, Appli

ALIGNMENTS

RESULT 1
US-08-357-598-1
; Sequence 1, Application US/08357598
; Patent No. 5705625
; GENERAL INFORMATION:
; APPLICANT: Civin, Curt I.
; APPLICANT: Small, Donald
; TITLE OF INVENTION: NOVEL PROTEIN TYROSINE KINASE, JAK3
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson P.C.
; STREET: 4225 Executive Square, Suite 1400
; CITY: La Jolla
; STATE: CA
; COUNTRY: USA
; ZIP: 92037

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/357,598
FILING DATE: 15-DEC-1994
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Halle, Lisa A.
REGISTRATION NUMBER: 38,347
REFERENCE/DOCKET NUMBER: 07265/033001
TELECOMMUNICATION INFORMATION:
TELEPHONE: 619/678-5070
TELEFAX: 619/678-5099
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 3807 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA
US-08-357-598-1

Query Match 69.2%; Score 16.6; DB 1; Length 3807;
Best Local Similarity 82.6%; Pred. No. 26;
Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 2 CTAGTCTGTCGAGGCATCTAGT 24

Db 1896 CAAGTCTGTCGAGGCATCTCGT 1918

```
RESULT 2
PCT-US95-16435-1
; Sequence 1, Application US/09003289
; Patent No. 5916792
; GENERAL INFORMATION:
; APPLICANT: Civin, Curt I.
; APPLICANT: Small, Donald
; TITLE OF INVENTION: NOVEL PROTEIN TYROSINE KINASE, JAK3
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson P.C.
; STREET: 4225 Executive Square, Suite 1400
; CITY: La Jolla
; STATE: CA
; COUNTRY: USA
; ZIP: 92037
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/003,289
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/357,598
; FILING DATE: 15-DEC-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Haile, Lisa A.
; REGISTRATION NUMBER: 38,347
; REFERENCE/DOCKET NUMBER: 07265/033001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 619/678-5099
; TELEFAX: 619/678-5099
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3807 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA
US-09-003-289-1
Query Match 69.2%; Score 16.6; DB 2; Length 3807;
Best Local Similarity 82.6%; Pred. No. 26;
Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 2 CTAGTGTCTGTCGAGGCATCTAGT 24
| | | | | | | | | | | | | | | | | |
Db 1896 CAAAGTGTCTGTCACCGCATCTCGT 1918

RESULT 3
PCT-US95-16435-1
; Sequence 1, Application PC/TUS9516435
; GENERAL INFORMATION:
; APPLICANT: The Johns Hopkins University School of Medicine
; TITLE OF INVENTION: NOVEL PROTEIN TYROSINE KINASE, JAK3
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson P.C.
; STREET: 4225 Executive Square, Suite 1400
; CITY: La Jolla
; STATE: CA
; COUNTRY: USA
; ZIP: 92037
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
```

```
; APPLICATION NUMBER: PCT/US95/16435
; FILING DATE: 15-DEC-1995
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Haile, Lisa A.
; REGISTRATION NUMBER: 38,347
; REFERENCE/DOCKET NUMBER: 07265/033001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 619/678-5070
; TELEFAX: 619/678-5099
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3807 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA
PCT-US95-16435-1
Query Match 69.2%; Score 16.6; DB 5; Length 3807;
Best Local Similarity 82.6%; Pred. No. 26;
Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 2 CTAGTGTCTGTCGAGGCATCTAGT 24
| | | | | | | | | | | | | | | | | |
Db 1896 CAAAGTGTCTGTCACCGCATCTCGT 1918

RESULT 4
US-08-250-848-2
; Sequence 2, Application US/08250848
; Patent No. 5856177
; GENERAL INFORMATION:
; APPLICANT: Hudspeth, Richard L.
; APPLICANT: Gula, John W.
; TITLE OF INVENTION: PHOTOMOTERS DERIVED FROM THE MAIZE
; TITLE OF INVENTION: PHOSPHOENOLPYRUVATE CARBOXYLASE GENE INVOLVED IN C4
; TITLE OF INVENTION: PHOTOSYNTHESIS
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Christie, Parker & Hale
; STREET: P.O. Box 7068
; CITY: Pasadena
; STATE: CA
; COUNTRY: USA
; ZIP: 91009-7068
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/250,848
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Sharp Esq., Janice A.
; REGISTRATION NUMBER: 34,051
; REFERENCE/DOCKET NUMBER: P114:25992/JAS
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (818) 795-5843
; TELEFAX: (818) 577-1769
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 7559 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Zea mays
```

```
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 2153..2332
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 2437..2832
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 2944..3027
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 3900..4121
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 4492..4596
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 4723..4812
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 4945..5100
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 5198..6196
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 6294..6680
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 6789..7079
; NAME/KEY: exon
; LOCATION: 2153..2332
; NAME/KEY: exon
; LOCATION: 2437..2832
; NAME/KEY: exon
; LOCATION: 2944..3027
; NAME/KEY: exon
; LOCATION: 3900..4121
; NAME/KEY: exon
; LOCATION: 4492..4596
; NAME/KEY: exon
; LOCATION: 4723..4812
; NAME/KEY: exon
; LOCATION: 4945..5100
; NAME/KEY: exon
; LOCATION: 5198..6196
; NAME/KEY: exon
; LOCATION: 6294..6680
; NAME/KEY: exon
; LOCATION: 6789..7079
; NAME/KEY: intron
; LOCATION: 2154..2331
; NAME/KEY: intron
; LOCATION: 2438..2831
; NAME/KEY: intron
; LOCATION: 2945..3026
; NAME/KEY: intron
; LOCATION: 3901..4120
; FEATURE:
```

```
; NAME/KEY: intron
; LOCATION: 4493..4595
; FEATURE:
; NAME/KEY: intron
; LOCATION: 4724..4811
; FEATURE:
; NAME/KEY: intron
; LOCATION: 4946..5099
; FEATURE:
; NAME/KEY: intron
; LOCATION: 5199..6195
; FEATURE:
; NAME/KEY: intron
; LOCATION: 6295..6679
; FEATURE:
; NAME/KEY: intron
; LOCATION: 6790..7078
; NAME/KEY: polyA_site
; LOCATION: 7314..7319
; FEATURE:
; NAME/KEY: prim_transcript
; LOCATION: 2072
; FEATURE:
; NAME/KEY: promoter
; LOCATION: 1..2155
; FEATURE:
; NAME/KEY: TATA_signal
; LOCATION: 2042..2049
; FEATURE:
; NAME/KEY: mRNA
; LOCATION: 2153..7079
; US-08-250-848-2
```

```
Query Match 69.2%; Score 16.6; DB 2; Length 7559;
Best Local Similarity 82.6%; Pred. No. 28;
Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
```

```
QY 2 CTAGTGTGTCGAGCATCTAGT 24
   ||| ||||| |||| ||| |||
Db 5176 CTACTGTGTCGATGTCAGT 5198
```

```
RESULT 5
US-09-433-579-3/c
; Sequence 3, Application US/09433579
; Patent No. 6444877
; GENERAL INFORMATION:
; APPLICANT: Rottmann, William H.
; TITLE OF INVENTION: LSAG Gene
; FILE REFERENCE: LSAG Gene
; CURRENT APPLICATION NUMBER: US/09/433,579
; CURRENT FILING DATE: 1999-11-04
; NUMBER OF SEQ ID NOS: 36
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 3
; LENGTH: 17656
; TYPE: DNA
; ORGANISM: Liquidambar styraciflua
; US-09-433-579-3
```

```
Query Match 66.7%; Score 16; DB 4; Length 17656;
Best Local Similarity 79.2%; Pred. No. 61;
Matches 19; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
```

```
QY 1 TCTAGTGTGTCGAGCATCTAGT 24
   ||||| |||| |||| |||
Db 15206 TCTAGTGGGTGCAAGCTGCTAGT 15183
```

```
RESULT 6
US-09-008-303-1/c
; Sequence 1, Application US/09008303
```

; Patent No. 6033889
; GENERAL INFORMATION:
; APPLICANT: HAN, YE SUN
; APPLICANT: YU, GYU YU
; APPLICANT: KIM, SUNG HOU
; APPLICANT: LIM, JAE HWAN
; APPLICANT: RYU, JAE RYEON
; APPLICANT: CHOI, IN GEOL
; TITLE OF INVENTION: GENE SEQUENCE OF AQUIFEX PYROPHILUS
; TITLE OF INVENTION: SUPEROXIDE DISMUTASE AND PROTEIN EXPRESSED IN ESCHERICHIA
; TITLE OF INVENTION: COLI
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, MAIER & NEUSTADT,
; ADDRESSEE: P.C.
; STREET: 1755 S. JEFFERSON DAVIS HIGHWAY, FOURTH FLOOR
; CITY: ARLINGTON
; STATE: VA
; COUNTRY: USA
; ZIP: 22202
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/008,303
; FILING DATE: 16-JAN-1998
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 97-1140
; FILING DATE: 16-JAN-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: OBLON, NORMAN F.
; REGISTRATION NUMBER: 24,618
; REFERENCE/DOCKET NUMBER: 2901-0109-0
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 703-413-3000
; TELEFAX: 703-413-2220
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2057 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1014..1652
; US-09-008-303-1

Query Match 64.2%; Score 15.4; DB 3; Length 2057;
Best Local Similarity 94.1%; Pred. No. 93;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 8 TCGTGCAGGCATCTAGT 24
|||||
DB 1362 TCGTGCAGGCATCAAGT 1346

RESULT 7
US-08-724-394A-20
; Sequence 20, Application US/08724394A
; Patent No. 5872237
; GENERAL INFORMATION:
; APPLICANT: Feder, John N.
; APPLICANT: Kronmal, Gregory S.
; APPLICANT: Lauer, Peter M.
; APPLICANT: Ruddy, David A.
; APPLICANT: Thomas, Winston
; APPLICANT: Tsuchihashi, Zenta
; APPLICANT: Wolff, Roger K.
; TITLE OF INVENTION: Megabase Transcript Map: No. 5872237el
; NUMBER OF SEQUENCES: 31
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: TOWNSEND and TOWNSEND and CREW LLP
; STREET: Two Embarcadero Center, 8th Floor
; CITY: San Francisco
; STATE: CA
; COUNTRY: USA
; ZIP: 94111-3834
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS

; TITLE OF INVENTION: Sequences and Antibodies Thereto
; NUMBER OF SEQUENCES: 31
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: TOWNSEND and TOWNSEND and CREW LLP
; STREET: Two Embarcadero Center, 8th Floor
; CITY: San Francisco
; STATE: CA
; COUNTRY: USA
; ZIP: 94111-3834
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/724,394A
; FILING DATE: 01-OCT-1996
; CLASSIFICATION: 536
; ATTORNEY/AGENT INFORMATION:
; NAME: Fitts, Renee A.
; REGISTRATION NUMBER: 35,136
; REFERENCE/DOCKET NUMBER: 017957-000100
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-576-0200
; TELEFAX: 415-576-0300
; INFORMATION FOR SEQ ID NO: 20:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 246240 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: not relevant
; TOPOLOGY: not relevant
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: misc-feature
; LOCATION: 1..246240
; OTHER INFORMATION: /note= "HLA-H.CONTIG"
; US-08-724-394A-20

Query Match 64.2%; Score 15.4; DB 2; Length 246240;
Best Local Similarity 94.1%; Pred. No. 1.6e+02;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 5 GTGTGTCGAGGCATCT 21
|||||
DB 142670 GTGTGTCGAGGCATCT 142686

RESULT 8
US-08-724-394A-21
; Sequence 21, Application US/08724394A
; Patent No. 5872237
; GENERAL INFORMATION:
; APPLICANT: Feder, John N.
; APPLICANT: Kronmal, Gregory S.
; APPLICANT: Lauer, Peter M.
; APPLICANT: Ruddy, David A.
; APPLICANT: Thomas, Winston
; APPLICANT: Tsuchihashi, Zenta
; APPLICANT: Wolff, Roger K.
; TITLE OF INVENTION: Megabase Transcript Map: No. 5872237el
; NUMBER OF SEQUENCES: 31
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: TOWNSEND and TOWNSEND and CREW LLP
; STREET: Two Embarcadero Center, 8th Floor
; CITY: San Francisco
; STATE: CA
; COUNTRY: USA
; ZIP: 94111-3834
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS

```
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/724,394A
; FILING DATE: 01-OCT-1996
; CLASSIFICATION: 536
; ATTORNEY/AGENT INFORMATION:
; NAME: Fitts, Renee A.
; REGISTRATION NUMBER: 35,136
; REFERENCE/DOCKET NUMBER: 017957-000100
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-576-0200
; TELEFAX: 415-576-0300
; INFORMATION FOR SEQ ID NO: 21:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 246240 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: not relevant
; TOPOLOGY: not relevant
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 1..246240
; OTHER INFORMATION: /note= "HLA-H.CONTIG"
; US-08-724-394A-21

Query Match 64.2%; Score 15.4; DB 2; Length 246240;
Best Local Similarity 94.1%; Pred. No. 1.6e+02;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 5 GTGTCGTGCAGGCATCT 21
|||||
Db 142670 GTGTGGTGCAGGCATCT 142686

RESULT 9
US-08-724-394A-22
; Sequence 22, Application US/08724394A
; Patent No. 5872237
; GENERAL INFORMATION:
; APPLICANT: Feder, John N.
; APPLICANT: Kronmal, Gregory S.
; APPLICANT: Lauer, Peter M.
; APPLICANT: Ruddy, David A.
; APPLICANT: Thomas, Winston
; APPLICANT: Tsuchihashi, Zenta
; APPLICANT: Wolff, Roger K.
; TITLE OF INVENTION: Megabase Transcript Map: No. 5872237el
; TITLE OF INVENTION: Sequences and Antibodies Thereto
; NUMBER OF SEQUENCES: 31
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: TOWNSEND and TOWNSEND and CREW LLP
; STREET: Two Embarcadero Center, 8th Floor
; CITY: San Francisco
; STATE: CA
; COUNTRY: USA
; ZIP: 94111-3834
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/724,394A
; FILING DATE: 01-OCT-1996
; CLASSIFICATION: 536
; ATTORNEY/AGENT INFORMATION:
; NAME: Fitts, Renee A.
; REGISTRATION NUMBER: 35,136
; REFERENCE/DOCKET NUMBER: 017957-000100
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-576-0200
; TELEFAX: 415-576-0300
; INFORMATION FOR SEQ ID NO: 22:
```

```
; SEQUENCE CHARACTERISTICS:
; LENGTH: 246240 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: not relevant
; TOPOLOGY: not relevant
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 1..246240
; OTHER INFORMATION: /note= "HLA-H.CONTIG"
; US-08-724-394A-22

Query Match 64.2%; Score 15.4; DB 2; Length 246240;
Best Local Similarity 94.1%; Pred. No. 1.6e+02;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 5 GTGTCGTGCAGGCATCT 21
|||||
Db 142670 GTGTGGTGCAGGCATCT 142686

RESULT 10
US-08-290-979A-7/c
; Sequence 7, Application US/08290979A
; Patent No. 5610046
; GENERAL INFORMATION:
; APPLICANT: VAN OOIJEN, Albert J.H.
; APPLICANT: DE GRAAFF, Leendert H.
; APPLICANT: VAN DEN BROECK, Henriette C.
; APPLICANT: VISSER, Jacob
; TITLE OF INVENTION: Cloning and Expression of Xylanase B
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MORRISON & FOERSTER
; STREET: 2000 Pennsylvania Ave. N.W.
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20006-1812
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/290,979A
; FILING DATE: 22-SEP-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: KATE H. MURASHIGE
; REGISTRATION NUMBER: 29,959
; REFERENCE/DOCKET NUMBER: 4615-0045.00
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 887-1500
; TELEFAX: (202) 887-0763
; TELEX: 90-4030 MRSN FOERS WSH
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2219 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ORIGINAL SOURCE:
; ORGANISM: Aspergillus tubigenis
; STRAIN: DSI6813
; FEATURE:
; NAME/KEY: CDS
; LOCATION: join(902..1180, 1248..1643)
; US-08-290-979A-7

Query Match 63.3%; Score 15.2; DB 1; Length 2219;
```

```
Best Local Similarity 85.0%; Pred. No. 1.2e+02;
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 TCTAGTGTGTCGAGGCATC 20
    ||||| ||| ||||| |||||
Db 624 TCTAGGTCTCTGCAGACATC 605

RESULT 11
US-09-328-111-155
; Sequence 155, Application US/09328111
; Patent No. 6262333
; GENERAL INFORMATION:
; APPLICANT: Endege, Wilson O.
; APPLICANT: Steinmann, Kathleen E.
; APPLICANT: Astle, Jon H.
; APPLICANT: Burgess, Christopher C.
; APPLICANT: Bushnell, Steven E.
; APPLICANT: Carroll III, Eddie
; APPLICANT: Catino, Theodore J.
; APPLICANT: Derti, Adnan
; APPLICANT: Ford, Donna M.
; APPLICANT: Lewis, Marcia E.
; APPLICANT: Monahan, John E.
; APPLICANT: Schlegel, Robert
; TITLE OF INVENTION: NOVEL HUMAN GENES AND GENE EXPRESSION
; TITLE OF INVENTION: PRODUCTS
; FILE REFERENCE: CCD-257 (US)
; CURRENT APPLICATION NUMBER: US/09/328,111
; CURRENT FILING DATE: 1999-06-08
; EARLIER APPLICATION NUMBER: US 60/088,801
; EARLIER FILING DATE: 1998-06-10
; NUMBER OF SEQ ID NOS: 850
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 155
; LENGTH: 594
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(594)
; OTHER INFORMATION: n = A,T,C or G
US-09-328-111-155

Query Match 62.5%; Score 15; DB 4; Length 594;
Best Local Similarity 75.0%; Pred. No. 1.3e+02;
Matches 18; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 1 TCTAGTGTGTCGAGGCATCTAGT 24
    || ||||| ||||| ||||| ||
Db 412 TCAAGTGTGTCGCTNCACTCCT 435

RESULT 12
US-08-448-744-6/C
; Sequence 6, Application US/08448744
; Patent No. 5616484
; GENERAL INFORMATION:
; APPLICANT: XU, Shuang-yong
; TITLE OF INVENTION: Cloning And Expression of The ApaLI
; TITLE OF INVENTION: Restriction Endonuclease
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: New England Biolabs, Inc.
; STREET: 32 Tozer Road
; CITY: Beverly
; STATE: Massachusetts
; COUNTRY: US
; ZIP: 01915
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS

Best Local Similarity 85.0%; Pred. No. 1.2e+02;
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 TCTAGTGTGTCGAGGCATC 20
    ||||| ||| ||||| |||||
Db 624 TCTAGGTCTCTGCAGACATC 605

SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/448,744
FILING DATE: 24-May-1995
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Williams, Gregory D.
REGISTRATION NUMBER: 30901
REFERENCE/DOCKET NUMBER: NEB-114
TELECOMMUNICATION INFORMATION:
TELEPHONE: (508) 927-5054
TELEFAX: (508) 927-1705
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 1128 base pairs
TYPE: nucleic acid
STRANDEDNESS: unknown
TOPOLOGY: unknown
MOLECULE TYPE: DNA (genomic)
US-08-448-744-6

Query Match 62.5%; Score 15; DB 1; Length 1128;
Best Local Similarity 78.3%; Pred. No. 1.4e+02;
Matches 18; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 TCTAGTGTGTCGAGGCATCTAG 23
    || ||||| ||||| ||||| ||
Db 302 TCCAGTCAAGTCGCTGCATCAAG 280

RESULT 13
US-09-245-041-3
; Sequence 3, Application US/09245041
; Patent No. 6274339
; GENERAL INFORMATION:
; APPLICANT: Moore, K.
; APPLICANT: Nagle, D.
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR THE DIAGNOSIS AND TREATMENT
; TITLE OF INVENTION: OF BODY WEIGHT DISORDERS INCLUDING OBESITY
; FILE REFERENCE: 7853-136
; CURRENT APPLICATION NUMBER: US/09/245,041
; CURRENT FILING DATE: 1999-02-05
; EARLIER APPLICATION NUMBER: 60/093,630
; EARLIER FILING DATE: 1998-07-21
; EARLIER APPLICATION NUMBER: 60/104,978
; EARLIER FILING DATE: 1998-10-20
; NUMBER OF SEQ ID NOS: 131
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 3
; LENGTH: 17056
; TYPE: DNA
; ORGANISM: Mus musculus
US-09-245-041-3

Query Match 62.5%; Score 15; DB 4; Length 17056;
Best Local Similarity 78.3%; Pred. No. 1.9e+02;
Matches 18; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 TCTAGTGTGTCGAGGCATCTAG 23
    || ||||| ||||| ||||| ||
Db 13217 TATAGTGTGTTCCAGGCATCCAG 13239

RESULT 14
US-07-644-372-1/C
; Sequence 1, Application US/07644372
; Patent No. 5416009
; GENERAL INFORMATION:
; APPLICANT: Lazzeri, Mario E.
; APPLICANT: Nutman, Thomas B.
; APPLICANT: Weiss, Niklaus
; TITLE OF INVENTION: A DNA SEGMENT ENCODING A SPECIFIC
; IMMUNODIAGNOSTIC ANTIGEN
```


NUMBER OF SEQUENCES: 2
CORRESPONDENCE ADDRESS:
ADDRESSEE: CUSHMAN, DARBY & CUSHMAN
STREET: 1615 L. Street, N.W.
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20036
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/644,372
FILING DATE: 19910123
CLASSIFICATION: 435
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202)861-3000
TELEFAX: (202)822-0944
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 822 base pairs
TYPE: NUCLEIC ACID
STRANDEDNESS: double
TOPOLOGY: linear
FEATURE:
NAME/KEY: CDS
LOCATION: 52...507
US-07-644-372-1

Query Match 61.7%; Score 14.8; DB 1; Length 822;
Best Local Similarity 88.9%; Pred. No. 1.6e+02;
Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
Qy 1 TCTAGTGTCTGTCAGGCA 18
Db 689 TCTAGTGTCTGTCAGCA 672

RESULT 15
US-08-296-014A-3
Sequence 3, Application US/08296014A
Patent No. 5716834
GENERAL INFORMATION:
APPLICANT: Ding, Jeak Ling
APPLICANT: Ho, Bow
TITLE OF INVENTION: The Cloned Factor C cDNA of the
TITLE OF INVENTION: Singapore Horseshoe Crab, Carcinoscorpius
TITLE OF INVENTION: rotundicauda and Purification of Factor C Proenzyme
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSEE: Birch, Stewart, Kolasch & Birch
STREET: 8110 Gatehouse Road, Suite 500 East
CITY: Falls Church
STATE: Virginia
COUNTRY: USA
ZIP: 22042
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/296,014A
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Murphy, Jr., Gerald M.
REGISTRATION NUMBER: 28,977
REFERENCE/DOCKET NUMBER: 1781-105P
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 203-8000

TELEFAX: (703) 203-8050
TELEX: 248345
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 3448 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: both
MOLECULE TYPE: cDNA
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Carcinoscorpius rotundicauda
FEATURE:
NAME/KEY: CDS
LOCATION: 18..3074
US-08-296-014A-3

Query Match 61.7%; Score 14.8; DB 1; Length 3448;
Best Local Similarity 88.9%; Pred. No. 2e+02;
Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
Qy 5 GTGTCGTGCGAGGCATCTA 22
Db 1136 GTGTCGTGCGAGGCATCCA 1153

Search completed: January 4, 2003, 00:11:20
Job time : 81.1009 secs

GenCore version 5.1.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 3, 2003, 22:54:41 ; Search time 8.10095 Seconds
(without alignments)
1281.345 Million cell updates/sec

Title: US-09-787-562-11

Perfect score: 24
Sequence: 1 tctagtgtgtgcaggcatctagt 24

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 381593 seqs, 216252194 residues

Total number of hits satisfying chosen parameters: 763186

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- Published Applications_NA:*
- 1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:*
 - 2: /cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq:*
 - 3: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq:*
 - 4: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq:*
 - 5: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq:*
 - 6: /cgn2_6/ptodata/1/pubpna/PCTUS_PUBCOMB.seq:*
 - 7: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq:*
 - 8: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:*
 - 9: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq:*
 - 10: /cgn2_6/ptodata/1/pubpna/US09_PUBCOMB.seq:*
 - 11: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq:*
 - 12: /cgn2_6/ptodata/1/pubpna/US10_PUBCOMB.seq:*
 - 13: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq:*
 - 14: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
C 1	16.6	69.2	1930	12	US-10-003-152-15
C 2	16.6	69.2	2056	12	US-10-003-152-23
C 3	16.2	67.5	371	10	US-09-960-352-2882
C 4	16.2	67.5	557	10	US-09-974-300-4795
C 5	16	66.7	159	10	US-09-864-761-22977
C 6	16	66.7	387	10	US-09-878-574-2017
C 7	16	66.7	456	10	US-09-864-761-6257
C 8	16	66.7	1005	10	US-09-815-242-6973
C 9	15.8	65.8	88191	10	US-09-799-799-3
C 10	15.6	65.0	200	10	US-09-864-761-24121
C 11	15.6	65.0	551	10	US-09-864-761-7458
C 12	15.6	65.0	1845	9	US-09-738-626-3487
C 13	15.6	65.0	40392	10	US-09-954-456-44
C 14	15.6	65.0	40392	10	US-09-954-456-44
C 15	15.4	64.2	598	9	US-10-001-835-62
C 16	15.2	63.3	299	10	US-09-294-093B-3520
C 17	15.2	63.3	684973	10	US-09-263-959-1
C 18	15	62.5	419	9	US-09-796-692-7030
C 19	15	62.5	594	10	US-09-879-536-155

C 20	15	62.5	954	10	US-09-815-242-9540	Sequence 9540, Ap
C 21	15	62.5	1197	9	US-09-938-842A-793	Sequence 793, App
C 22	15	62.5	2141	10	US-09-917-800A-1511	Sequence 1511, Ap
C 23	15	62.5	2256	10	US-09-899-471-1	Sequence 1, Appli
C 24	15	62.5	2328	10	US-09-899-471-4	Sequence 4, Appli
C 25	15	62.5	2538	9	US-10-098-841-313	Sequence 313, App
C 26	15	62.5	17056	10	US-09-893-238-3	Sequence 3, Appli
C 27	15	62.5	167343	10	US-09-962-436-281	Sequence 281, App
C 28	15	62.5	167343	10	US-09-964-824A-273	Sequence 273, App
C 29	15	62.5	174424	10	US-09-967-768A-314	Sequence 314, App
C 30	15	62.5	198285	10	US-09-880-107-3814	Sequence 3814, Ap
C 31	14.8	61.7	217	10	US-09-867-701-4514	Sequence 4514, Ap
C 32	14.8	61.7	580	10	US-09-864-761-8242	Sequence 8242, Ap
C 33	14.8	61.7	4369	10	US-09-769-097-1	Sequence 1, Appli
C 34	14.8	61.7	4425	10	US-09-769-097-3	Sequence 3, Appli
C 35	14.8	61.7	15535	10	US-09-764-877-2855	Sequence 2855, Ap
C 36	14.8	61.7	203654	10	US-09-820-905-3	Sequence 3, Appli
C 37	14.6	60.8	127	10	US-09-867-701-10523	Sequence 10523, A
C 38	14.6	60.8	192	10	US-09-867-701-8124	Sequence 8124, Ap
C 39	14.6	60.8	196	10	US-09-864-761-30975	Sequence 30975, A
C 40	14.6	60.8	374	9	US-09-886-242A-3	Sequence 3, Appli
C 41	14.6	60.8	374	9	US-10-027-603-3	Sequence 3, Appli
C 42	14.6	60.8	387	10	US-09-983-965-3685	Sequence 3685, Ap
C 43	14.6	60.8	429	9	US-09-796-692-8460	Sequence 8460, Ap
C 44	14.6	60.8	457	10	US-09-864-761-14419	Sequence 14419, A
C 45	14.6	60.8	461	10	US-09-864-761-15009	Sequence 15009, A

ALIGNMENTS

RESULT 1
US-10-003-152-15/c
; Sequence 15, Application US/10003152
; Patent No. US20020151494A1
; GENERAL INFORMATION:
; APPLICANT: Shimkets, Richard
; APPLICANT: Fernandes, Elma
; APPLICANT: Vernet, Corine
; APPLICANT: Yang, Meljia
; APPLICANT: Boldog, Ferenc
; APPLICANT: Herimann, John
; TITLE OF INVENTION: No. US20020151494A1e1 Amino Acid Sequences for Human Semaphori
; FILE REFERENCE: 15966-554 Cura-54 CON-S12
; CURRENT APPLICATION NUMBER: US/10/003,152
; CURRENT FILING DATE: 2001-11-02
; PRIOR APPLICATION NUMBER: 09/604,286
; PRIOR FILING DATE: 2000-06-22
; PRIOR APPLICATION NUMBER: 60/140,584
; PRIOR FILING DATE: 1999-06-23
; NUMBER OF SEQ ID NOS: 49
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 15
; LENGTH: 1930
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (148)..(1755)
; NAME/KEY: variation
; LOCATION: (1)..(1930)
; OTHER INFORMATION: N may be any nucleotide
US-10-003-152-15

Query Match 69.2%; Score 16.6; DB 12; Length 1930;
Best Local Similarity 82.6%; Pred. No. 36;
Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
QY 2 CTAGTGTCTGTCAGGCATCTAGT 24
Db 284 CTAGTCTCGTCAGGCATCTCGT 262

```
RESULT 2
US-10-003-152-23/C
; Sequence 23, Application US/10003152
; Patent No. US20020151494A1
; GENERAL INFORMATION:
; APPLICANT: Shimkets, Richard
; APPLICANT: Fernandes, Elma
; APPLICANT: Vernet, Corine
; APPLICANT: Yang, Meijia
; APPLICANT: Boldog, Ferenc
; APPLICANT: Herrmann, John
; TITLE OF INVENTION: NO. US20020151494A1el Amino Acid Sequences for Human Semaphorin-1
; FILE REFERENCE: 15966-554 Cura-54 CON-S12
; CURRENT APPLICATION NUMBER: US/10/003.152
; CURRENT FILING DATE: 2001-11-02
; PRIOR APPLICATION NUMBER: 09/604,286
; PRIOR FILING DATE: 2000-06-22
; PRIOR APPLICATION NUMBER: 60/140,584
; PRIOR FILING DATE: 1999-06-23
; NUMBER OF SEQ ID NOS: 49
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 23
; LENGTH: 2056
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (7)..(1608)
; NAME/KEY: variation
; LOCATION: (1)..(2056)
; OTHER INFORMATION: N may be any nucleotide
; NAME/KEY: variation
; LOCATION: (1)..(2056)
; OTHER INFORMATION: N may be any nucleotide
US-10-003-152-23

Query Match 69.2%; Score 16.6; DB 12; Length 2056;
Best Local Similarity 82.6%; Pred. No. 36;
Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 2 CTAGTGTGTCGAGGCATCTAGT 24
||||| ||||||| |||||||
Db 137 CTAGTCTCGTCGACGGATCTCGT 115

RESULT 3
US-09-960-352-2882
; Sequence 2882, Application US/09960352
; Patent No. US20020137139A1
; GENERAL INFORMATION:
; APPLICANT: Warren, Wesley C.
; APPLICANT: Tao, Nengbing
; APPLICANT: Byatt, John C.
; APPLICANT: Mathialagan, Nagappan
; TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND
; FILE REFERENCE: 16511.006/37-21(10298)C
; CURRENT APPLICATION NUMBER: US/09/960,352
; CURRENT FILING DATE: 2001-09-24
; NUMBER OF SEQ ID NOS: 15112
; SEQ ID NO 2882
; LENGTH: 371
; TYPE: DNA
; ORGANISM: Bos taurus
; OTHER INFORMATION: Clone ID: 13-LIB3057-007-Q1-K1-D1
US-09-960-352-2882

Query Match 67.5%; Score 16.2; DB 10; Length 371;
Best Local Similarity 85.7%; Pred. No. 50;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 3 TAGTGTGTCGACGGATCTAGT 23
| ||||| ||||||| |||||||
```

```
Db 224 TTGTGTCATGCAGGCATCTAG 244

RESULT 4
US-09-974-300-4795
; Sequence 4795, Application US/09974300
; Patent No. US20020146721A1
; GENERAL INFORMATION:
; APPLICANT: Berka, Randy M.
; APPLICANT: Clausen, ID Groth
; TITLE OF INVENTION: Methods for Monitoring Multiple Gene
; FILE REFERENCE: 10085.500-US
; CURRENT APPLICATION NUMBER: US/09/974,300
; CURRENT FILING DATE: 2001-10-05
; PRIOR APPLICATION NUMBER: 09/680,598
; PRIOR FILING DATE: 2000-10-06
; PRIOR APPLICATION NUMBER: 60/279,526
; PRIOR FILING DATE: 2001-03-27
; NUMBER OF SEQ ID NOS: 8481
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4795
; LENGTH: 557
; TYPE: DNA
; ORGANISM: Bacillus clausii
; FEATURE:
; NAME/KEY: misc.feature
; LOCATION: (1)..(557)
; OTHER INFORMATION: n = A,T,C or G
US-09-974-300-4795

Query Match 67.5%; Score 16.2; DB 10; Length 557;
Best Local Similarity 85.7%; Pred. No. 52;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 CTAGTGTGTCGAGGCATCTA 22
||||| ||||| ||||| |||||
Db 292 CTCGTGTCCTGCAAGCATCTA 312

RESULT 5
US-09-864-761-22977/c
; Sequence 22977, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharron G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FO
; FILE REFERENCE: Aeomica-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
```

;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: PCT/US01/00668
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: PCT/US01/00663
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: PCT/US01/00662
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: PCT/US01/00661
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: PCT/US01/00670
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: US 60/234,687
;; PRIOR FILING DATE: 2000-09-21
;; PRIOR APPLICATION NUMBER: US 09/608,408
;; PRIOR FILING DATE: 2000-06-30
;; PRIOR APPLICATION NUMBER: US 09/774,203
;; PRIOR FILING DATE: 2001-01-29
;; NUMBER OF SEQ ID NOS: 49117
;; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
;; SEQ ID NO 22977
;; LENGTH: 159
;; TYPE: DNA
;; ORGANISM: Homo sapiens
;; FEATURE:
;; OTHER INFORMATION: MAP TO AC004549.1
;; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 15
;; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 13
;; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 15
;; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 14
;; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 15
;; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 17
;; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 14
;; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 15
;; OTHER INFORMATION: NT HIT: AF135399.1, EVALUATE 3.60e-01
;; OTHER INFORMATION: EST_HUMAN HIT: AI833003.1, EVALUATE 2.00e-77

US-09-864-761-22977

Query Match 66.7%; Score 16; DB 10; Length 159;
Best Local Similarity 79.2%; Pred. No. 60;
Matches 19; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 TCTAGTGTGTCGACGCACTAGT 24
||| ||||| ||||| ||||| |||||
DB 158 TCTCTGTGTCGACGCACTAGT 135

RESULT 6

US-09-878-574-2017
;; Sequence 2017, Application US/09878574
;; Patent No. US20020110548A1
;; GENERAL INFORMATION:
;; APPLICANT: Byrum, Joseph R.
;; APPLICANT: La Rosa, Thomas J.
;; APPLICANT: Thompson, Michael D.
;; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated with
;; TITLE OF INVENTION: Plants
;; FILE REFERENCE: 38-21(15401)B
;; CURRENT APPLICATION NUMBER: US/09/878,574
;; CURRENT FILING DATE: 2001-12-21
;; PRIOR APPLICATION NUMBER: 09/333,535
;; PRIOR FILING DATE: 1999-06-14
;; NUMBER OF SEQ ID NOS: 15775
;; SEQ ID NO 2017
;; LENGTH: 387
;; TYPE: DNA
;; ORGANISM: Glycine max
;; OTHER INFORMATION: Clone ID: LIB3028-030-Q1-B1-C4
US-09-878-574-2017

Query Match 66.7%; Score 16; DB 10; Length 387;
Best Local Similarity 79.2%; Pred. No. 63;
Matches 19; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 TCTAGTGTGTCGACGCACTAGT 24
||| ||||| ||||| ||||| |||||
DB 258 TCTACACTCTTGCAGCCATCTAGT 281

RESULT 7

US-09-864-761-6257/c
;; Sequence 6257, Application US/09864761
;; Patent No. US20020048763A1
;; GENERAL INFORMATION:
;; APPLICANT: Penn, Sharron G.
;; APPLICANT: Rank, David R.
;; APPLICANT: Hanzel, David K.
;; APPLICANT: Chen, Wensheng
;; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
;; TITLE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY
;; FILE REFERENCE: Acomica-X-1
;; CURRENT APPLICATION NUMBER: US/09/864,761
;; CURRENT FILING DATE: 2001-05-23
;; PRIOR APPLICATION NUMBER: US 60/180,312
;; PRIOR FILING DATE: 2000-02-04
;; PRIOR APPLICATION NUMBER: US 60/207,456
;; PRIOR FILING DATE: 2000-05-26
;; PRIOR APPLICATION NUMBER: US 09/632,366
;; PRIOR FILING DATE: 2000-08-03
;; PRIOR APPLICATION NUMBER: GB 24263.6
;; PRIOR FILING DATE: 2000-10-04
;; PRIOR APPLICATION NUMBER: US 60/236,359
;; PRIOR FILING DATE: 2000-09-27
;; PRIOR APPLICATION NUMBER: PCT/US01/00666
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: PCT/US01/00667
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: PCT/US01/00664
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: PCT/US01/00669
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: PCT/US01/00665
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: PCT/US01/00668
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: PCT/US01/00663
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: PCT/US01/00662
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: PCT/US01/00661
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: PCT/US01/00670
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: US 60/234,687
;; PRIOR FILING DATE: 2000-09-21
;; PRIOR APPLICATION NUMBER: US 09/608,408
;; PRIOR FILING DATE: 2000-06-30
;; PRIOR APPLICATION NUMBER: US 09/774,203
;; PRIOR FILING DATE: 2001-01-29
;; NUMBER OF SEQ ID NOS: 49117
;; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
;; SEQ ID NO 6257
;; LENGTH: 456
;; TYPE: DNA
;; ORGANISM: Homo sapiens
;; FEATURE:
;; OTHER INFORMATION: MAP TO AC004549.1
;; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 15
;; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 13
;; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 15
;; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 15
;; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 14
;; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 17
;; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 14
;; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 15
US-09-864-761-6257

Query Match 66.7%; Score 16; DB 10; Length 456;
Best Local Similarity 79.2%; Pred. No. 64;
Matches 19; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 1 TCTAGTGTCTGCAGGCATCTAGT 24
||| |||| |||| |||| |||| |||| |||| |||| ||||
Db 435 TCTCTTGTCTGCAGGCAGCGAGT 412

RESULT 8

US-09-815-242-6973

; Sequence 6973, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Karl L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; TITLE OF INVENTION: Prokaryotes
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 6973
; LENGTH: 1005
; TYPE: DNA
; ORGANISM: Haemophilus influenzae
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)...(1005)
US-09-815-242-6973

Query Match 66.7%; Score 16; DB 10; Length 1005;
Best Local Similarity 79.2%; Pred. No. 67;
Matches 19; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 1 TCTAGTGTCTGCAGGCATCTAGT 24
||| |||| |||| |||| |||| |||| |||| |||| ||||
Db 19 TTAGTGGCGTGCAGCCTCTCGT 42

RESULT 9

US-09-799-799-3

; Sequence 3, Application US/09799799
; Patent No. US20020132291A1
; GENERAL INFORMATION:
; APPLICANT: YE, Jane et al.
; TITLE OF INVENTION: ISOLATED HUMAN RAS-LIKE PROTEINS,
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING THESE HUMAN RAS-LIKE
; TITLE OF INVENTION: PROTEINS, AND USES THEREOF
; FILE REFERENCE: C1001157
; CURRENT APPLICATION NUMBER: US/09/799,799
; CURRENT FILING DATE: 2001-03-07

; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 88191
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(88191)
; OTHER INFORMATION: n = A,T,C or G
US-09-799-799-3

Query Match 65.8%; Score 15.8; DB 10; Length 88191;
Best Local Similarity 89.5%; Pred. No. 1.1e+02;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 6 TGTCTGTGCAGGCATCTAGT 24
|||| |||| |||| |||| |||| |||| |||| |||| ||||
Db 72640 TGTCTGTGCAGGCTTCTAGT 72658

RESULT 10

US-09-864-761-24121
; Sequence 24121, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharron G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; TITLE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY
; FILE REFERENCE: Aecmca-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1


```
US-09-954-456-44
; Sequence 44, Application US/09954456
; Patent No. US20020115057A1
; GENERAL INFORMATION:
; APPLICANT: Young, Paul
; TITLE OF INVENTION: Process for Identifying Anti-Cancer Therapeutic Agents Using Cand
; FILE OF INVENTION: Sets
; FILE REFERENCE: 689290-76
; CURRENT APPLICATION NUMBER: US/09/954,456
; PRIOR FILING DATE: 2001-09-18
; PRIOR APPLICATION NUMBER: US/60/233,617
; PRIOR FILING DATE: 2000-09-18
; PRIOR APPLICATION NUMBER: US/60/234,052
; PRIOR FILING DATE: 2000-09-20
; PRIOR APPLICATION NUMBER: US/60/234,923
; PRIOR FILING DATE: 2000-09-25
; PRIOR APPLICATION NUMBER: US/60/235,134
; PRIOR FILING DATE: 2000-09-25
; PRIOR APPLICATION NUMBER: US/60/235,637
; PRIOR FILING DATE: 2000-09-26
; PRIOR APPLICATION NUMBER: US/60/235,638
; PRIOR FILING DATE: 2000-09-26
; PRIOR APPLICATION NUMBER: US/60/235,711
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: US/60/235,720
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: US/60/235,840
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: US/60/235,863
; PRIOR FILING DATE: 2000-09-27
; NUMBER OF SEQ ID NOS: 2276
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 44
; LENGTH: 40392
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-954-456-44

Query Match 65.0%; Score 15.6; DB 10; Length 40392;
Best Local Similarity 81.8%; Pred. No. 1.3e+02;
Matches 18; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 3 TAGTGTCTGTCAGGCATCTAGT 24
|| |||| |||| |||| |||| ||
Db 27507 TATTCTCTGCAAGCATCTGGT 27528

RESULT 14
US-09-954-456-687
; Sequence 687, Application US/09954456
; Patent No. US20020115057A1
; GENERAL INFORMATION:
; APPLICANT: Young, Paul
; TITLE OF INVENTION: Process for Identifying Anti-Cancer Therapeutic Agents Using Cand
; FILE OF INVENTION: Sets
; FILE REFERENCE: 689290-76
; CURRENT APPLICATION NUMBER: US/09/954,456
; PRIOR FILING DATE: 2001-09-18
; PRIOR APPLICATION NUMBER: US/60/233,617
; PRIOR FILING DATE: 2000-09-18
; PRIOR APPLICATION NUMBER: US/60/234,052
; PRIOR FILING DATE: 2000-09-20
; PRIOR APPLICATION NUMBER: US/60/235,134
; PRIOR FILING DATE: 2000-09-25
; PRIOR APPLICATION NUMBER: US/60/235,637
; PRIOR FILING DATE: 2000-09-26
; PRIOR APPLICATION NUMBER: US/60/235,638
; PRIOR FILING DATE: 2000-09-26
; PRIOR APPLICATION NUMBER: US/60/235,711
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: US/60/235,720
; NUMBER OF SEQ ID NOS: 2276
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 44
; LENGTH: 40392
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-954-456-44

Query Match 65.0%; Score 15.6; DB 10; Length 40392;
Best Local Similarity 81.8%; Pred. No. 1.3e+02;
Matches 18; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 3 TAGTGTCTGTCAGGCATCTAGT 24
|| |||| |||| |||| |||| ||
Db 27507 TATTCTCTGCAAGCATCTGGT 27528

RESULT 15
US-10-001-835-62/c
; Sequence 62, Application US/10001835
; Patent No. US20020160387A1
; GENERAL INFORMATION:
; APPLICANT: Salceda, Susana
; APPLICANT: Macina, Roberto
; APPLICANT: Recipon, Herve
; APPLICANT: Cafferkey, Robert
; APPLICANT: Sun, Yongming
; APPLICANT: Liu, Chenghua
; TITLE OF INVENTION: Compositions and Methods Relating to Ovary Specific Genes and
; FILE REFERENCE: DEX-0277
; CURRENT APPLICATION NUMBER: US/10/001,835
; CURRENT FILING DATE: 2001-11-20
; PRIOR APPLICATION NUMBER: 60/249,997
; PRIOR FILING DATE: 2000-11-20
; NUMBER OF SEQ ID NOS: 228
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 62
; LENGTH: 598
; TYPE: DNA
; ORGANISM: Homo sapien
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (270)..(270)
; OTHER INFORMATION: a, c, g or t
; NAME/KEY: misc_feature
; LOCATION: (330)..(330)
; OTHER INFORMATION: a, c, g or t
; NAME/KEY: misc_feature
; LOCATION: (332)..(332)
; OTHER INFORMATION: a, c, g or t
; NAME/KEY: misc_feature
; LOCATION: (334)..(334)
; OTHER INFORMATION: a, c, g or t
; NAME/KEY: misc_feature
; LOCATION: (533)..(533)
; OTHER INFORMATION: a, c, g or t
US-10-001-835-62

Query Match 64.2%; Score 15.4; DB 9; Length 598;
Best Local Similarity 94.1%; Pred. No. 1.3e+02;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 8 TCGTGCAGGCATCTAGT 24
|| || || || || || || || || ||
Db 59 TCGAGCAGGCATCTAGT 43
```

Search completed: January 4, 2003, 01:06:35

• Mon Jan 6 15:20:21 2003

us-09-787-562-11.rnpb

Page 7

Job time : 25.1009 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 3, 2003, 22:52:46 ; Search time 2134.5 Seconds
(without alignments)
3231.380 Million cell updates/sec

Title: US-09-787-562-9
Perfect score: 237
Sequence: 1 gctagagtcgtgcagagcgt.....cgaggccgcctggcctctg 237

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues
Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

GenEmbl.*

- 1: gb.ba.*
- 2: gb.htg.*
- 3: gb.in.*
- 4: gb.om.*
- 5: gb.ov.*
- 6: gb.pat.*
- 7: gb.ph.*
- 8: gb.pl.*
- 9: gb.pr.*
- 10: gb.ro.*
- 11: gb.sts.*
- 12: gb.sy.*
- 13: gb.un.*
- 14: gb.vi.*
- 15: em.ba.*
- 16: em.fun.*
- 17: em.hum.*
- 18: em.in.*
- 19: em.mu.*
- 20: em.om.*
- 21: em.or.*
- 22: em.ov.*
- 23: em.pat.*
- 24: em.ph.*
- 25: em.pl.*
- 26: em.ro.*
- 27: em.sts.*
- 28: em.un.*
- 29: em.vi.*
- 30: em.htg_hum.*
- 31: em.htg_inv.*
- 32: em.htg_other.*
- 33: em.htg_mus.*
- 34: em.htg_pln.*
- 35: em.htg_fod.*
- 36: em.htg_mam.*
- 37: em.htg_vrt.*
- 38: em.sy.*
- 39: em.htgo_hum.*
- 40: em.htgo_mus.*
- 41: em.htgo_other.*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	237	100.0	237	6	AX023667	AX023667 Sequence
2	221	93.2	243	6	AX018563	AX018563 Sequence
3	221	93.2	243	6	AX018639	AX018639 Sequence
4	221	93.2	243	6	AX023661	AX023661 Sequence
5	175	73.8	242	6	AX023682	AX023682 Sequence
6	171	72.2	4219	12	RVU57027	U57027 Reporter vec
7	171	72.2	4465	12	RVU57025	U57025 Reporter vec
8	171	72.2	4869	12	CVU89940	U89940 Cloning vec
9	171	72.2	5010	6	AX339210	AX339210 Sequence
10	171	72.2	5010	12	CVU47298	U47298 Cloning vec
11	171	72.2	5115	12	CVU89938	U89938 Cloning vec
12	171	72.2	5256	6	AX339208	AX339208 Sequence
13	171	72.2	5256	12	CVU47296	U47296 Cloning vec
14	160	67.5	6320	12	AB038600	AB038600 Cloning v
15	157.2	66.3	5917	12	AB037684	AB037684 Cloning v
16	152.2	64.2	267	6	AX023685	AX023685 Sequence
17	148.4	62.6	204	6	AX023687	AX023687 Sequence
18	148	62.4	223	6	AX023683	AX023683 Sequence
19	147.4	62.2	356	6	AS8090	AS8090 Sequence 1
20	147.4	62.2	356	6	AR095774	AR095774 Sequence
21	146.8	61.9	5789	12	CVU962PRO	X65326 Cloning vec
22	146.8	61.9	5991	12	CVU09663	X09663 Cloning vec
23	146.8	61.9	6046	12	CVU09663	X65324 Cloning vec
24	146.8	61.9	6248	12	CVU09661	X09661 Cloning vec
25	146.8	61.9	7674	12	CVU13187	U13187 Cloning vec
26	146.8	61.9	7931	12	CVU13186	U13186 Cloning vec
27	146.4	61.8	4506	12	CVU096PRO	X65320 Cloning vec
28	146.4	61.8	4750	12	CVU096CON	X65321 Cloning vec
29	145.2	61.3	7731	6	A94048	A94048 sequence 29
30	145.2	61.3	7731	6	A94061	A94061 sequence 42
31	145.2	61.3	7731	6	AX011133	AX011133 Sequence
32	145.2	61.3	7731	6	AX011146	AX011146 Sequence
33	144.8	61.1	5894	12	AF334827	AF334827 Cloning v
34	144.4	60.9	8772	12	AF136442	AF136442 Cloning v
35	143.6	60.6	110000	2	LMFLCHR34_08	Continuation (9 of
36	142.8	60.3	350	6	E00567	E00567 DNA sequenc
37	142.8	60.3	455	6	E01399	E01399 DNA sequenc
38	142.8	60.3	5003	12	SYNPSV2CAT	M77788 pSV2-Cat cl
39	142.8	60.3	5840	12	NEIEXPVECB	L07041 pMNEO euka
40	142.8	60.3	5846	12	NEIEXPVECA	L07040 pFNEO euka
41	142.8	60.3	8068	6	A94046	A94046 Sequence 27
42	142.8	60.3	8068	6	A94054	A94054 Sequence 35
43	142.8	60.3	8068	6	AX011131	AX011131 Sequence
44	142.8	60.3	8068	6	AX011139	AX011139 Sequence
45	142.8	60.3	13254	6	AR038307	AR038307 Sequence

ALIGNMENTS

RESULT	1	AX023667	Sequence 9 from Patent WO0017371.	237 bp	DNA	linear	PAT 15-SEP-2000
LOCUS	AX023667	Sequence 9 from Patent WO0017371.	237 bp	DNA	linear	PAT 15-SEP-2000	
DEFINITION	AX023667	Sequence 9 from Patent WO0017371.	237 bp	DNA	linear	PAT 15-SEP-2000	
ACCESSION	AX023667	Sequence 9 from Patent WO0017371.	237 bp	DNA	linear	PAT 15-SEP-2000	
VERSION	AX023667.1	GI:10184028					
KEYWORDS		synthetic construct.					
SOURCE		synthetic construct.					
ORGANISM		artificial sequences.					
REFERENCE		1 (bases 1 to 237)					
AUTHORS		Binley,K.M. and Naylor,S.					
TITLE		Polynucleotide constructs and uses thereof					
JOURNAL		Patent: WO 0017371-A 9 30-MAR-2000;					
		BINLEY KATIE MARY (GB) ; NAYLOR STUART (GB) ; OXFORD BIOMEDICA LTD					

(GB) Location/Qualifiers
 1..237
 /organism="synthetic construct"
 /db_xref="taxon:32630"
 /note="OBHrelli"
 43 a 82 c 56 g 56 t
 BASE COUNT
 ORIGIN
 Query Match 100.0%; Score 237; DB 6; Length 237;
 Best Local Similarity 100.0%; Pred. No. 1.6e-62;
 Matches 237; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 GCTAGAGTGTGCAGGACGTGACATCTAGTGTGCGGAGGATCTAGTGTGCGTGCAGGAC 60
 DB 1 GCTAGAGTGTGCAGGACGTGACATCTAGTGTGCGGAGGATCTAGTGTGCGTGCAGGAC 60
 QY 61 GTGACAGCTAGCCGGGCTGAGATCTGGATCTGCATCTCAATTAGTGCACACCATAG 120
 DB 61 GTGACAGCTAGCCGGGCTGAGATCTGGATCTGCATCTCAATTAGTGCACACCATAG 120
 QY 121 TCCGGCCCTAACTCCGCCATCCGCCCTAACTCCGCCAGTTCGCCCATCTCCGC 180
 DB 121 TCCGGCCCTAACTCCGCCATCCGCCCTAACTCCGCCAGTTCGCCCATCTCCGC 180
 QY 181 CCCATGCTGACTAAATTTTTTTTATTTATGACAGGCGGAGCCGCTCGGCCTCTG 237
 DB 181 CCCATGCTGACTAAATTTTTTTTATTTATGACAGGCGGAGCCGCTCGGCCTCTG 237
 RESULT 2
 AX018563
 LOCUS AX018563 243 bp DNA linear PAT 07-SEP-2000
 DEFINITION Sequence 57 from Patent WO9945127.
 ACCESSION AX018563
 VERSION AX018563.1 GI:10042701
 KEYWORDS
 SOURCE synthetic construct.
 ORGANISM
 REFERENCE 1 (bases 1 to 243)
 AUTHORS Kingsman,S.M., Miropahanous,K., Patterson,A.V., Stratford,I.J.,
 Griffiths,L. and Kan,O.
 TITLE Enhanced produg activation
 JOURNAL Patent: WO 9945127-A 57 10-SEP-1999;
 KINGSMAN SUSAN MARY (GB); MIROPHANOUS KYRIACOS (GB); PATTERSON
 ADAM VORN (GB); STRATFORD IAN JAMES (GB); GRIFFITHS LEIGH (GB); KAN
 ON (GB); OXFORD BIOMEDICA LTD (GB)
 FEATURES
 source
 1..243
 /organism="synthetic construct"
 /db_xref="taxon:32630"
 /note="Promoter"
 45 a 83 c 58 g 57 t
 BASE COUNT
 ORIGIN
 Query Match 93.2%; Score 221; DB 6; Length 243;
 Best Local Similarity 97.5%; Pred. No. 1.4e-57;
 Matches 237; Conservative 0; Mismatches 0; Indels 6; Gaps 1;
 QY 1 GCTAGAGTGTGCGGAGGCGTGCATCTAGTGTGCGTGCAGG-----CATCTAGTGTGCTG 54
 DB 1 GCTAGAGTGTGCGGAGGCGTGCATCTAGTGTGCGTGCAGGAGCGTGACATCTAGTGTGCTG 60
 QY 55 CAGGACGTGACAGCTAGCCGGCTCGAGATCTGCATCTCAATTAGTGCACAA 114
 DB 61 CAGGACGTGACAGCTAGCCGGCTCGAGATCTGCATCTCAATTAGTGCACAA 120
 QY 115 CCATAGTCCCGCCCTAACTCCGCCATCCGCCCTAACTCCGCCAGTTCGCCCATTT 174
 DB 121 CCATAGTCCCGCCCTAACTCCGCCATCCGCCCTAACTCCGCCAGTTCGCCCATTT 180
 QY 175 CTCGCCGCCATCGCTGACTAAATTTTTTTTATTTATGACAGGCGGAGCCGCTCGGCCT 234
 DB 181 CTCGCCGCCATCGCTGACTAAATTTTTTTTATTTATGACAGGCGGAGCCGCTCGGCCT 240
 QY 235 CTG 237
 DB 241 CTG 243
 RESULT 4
 AX023661
 LOCUS AX023661 243 bp DNA linear PAT 15-SEP-2000
 DEFINITION Sequence 3 from Patent WO0017371.
 ACCESSION AX023661
 VERSION AX023661.1 GI:10184022
 KEYWORDS
 SOURCE synthetic construct.
 ORGANISM
 REFERENCE 1 (bases 1 to 243)
 AUTHORS Binley,K.M. and Naylor,S.
 TITLE Polynucleotide constructs and uses thereof
 JOURNAL Patent: WO 0017371-A 3 30-MAR-2000;
 QY 1 GCTAGAGTGTGCGGAGGCGTGCATCTAGTGTGCGTGCAGG-----CATCTAGTGTGCTG 54
 DB 1 GCTAGAGTGTGCGGAGGCGTGCATCTAGTGTGCGTGCAGGAGCGTGACATCTAGTGTGCTG 60
 QY 55 CAGGACGTGACAGCTAGCCGGCTCGAGATCTGCATCTCAATTAGTGCACAA 114
 DB 61 CAGGACGTGACAGCTAGCCGGCTCGAGATCTGCATCTCAATTAGTGCACAA 120
 QY 115 CCATAGTCCCGCCCTAACTCCGCCATCCGCCCTAACTCCGCCAGTTCGCCCATTT 174
 DB 121 CCATAGTCCCGCCCTAACTCCGCCATCCGCCCTAACTCCGCCAGTTCGCCCATTT 180
 QY 175 CTCGCCGCCATCGCTGACTAAATTTTTTTTATTTATGACAGGCGGAGCCGCTCGGCCT 234

BINLEY KATIE MARY (GB) ; NAYLOR STUART (GB) ; OXFORD BIOMEDICA LTD (GB)

FEATURES

Location/Qualifiers

1..243

/organism="synthetic construct"

/db_xref="taxon:32630"

/note="OBHrel"

BASE COUNT 45 a 83 c 58 g 57 t

ORIGIN

Query Match 93.2%; Score 221; DB 6; Length 243;

Best Local Similarity 97.5%; Pred. No. 1.4e-57;

Matches 237; Conservative 0; Mismatches 0; Indels 6; Gaps 1;

QY 1 GCTAGAGTCGTCAGAGCGTACATCTAGTCTGCTGCGAGG-----CATCTAGTGTCTGTG 54

Db 1 GCTAGAGTCGTCAGAGCGTACATCTAGTCTGCTGCGAGCGTGCATCTAGTGTCTGTG 60

QY 55 CAGGACGTGACAGCTAGCGCGGCTCGAGATCTGCGATCTGCATCTCAATTAGTCAGCAA 114

Db 61 CAGGACGTGACAGCTAGCGCGGCTCGAGATCTGCGATCTGCATCTCAATTAGTCAGCAA 120

QY 115 CCATAGTCCCGCCCTAATCTCGCCCATCCCGCCCTAACTCCGCCAGTTCCCGCCCAT 174

Db 121 CCATAGTCCCGCCCTAATCTCGCCCATCCCGCCCTAACTCCGCCAGTTCCCGCCCAT 180

QY 175 CTCGCCCATCTCGCTGACTAATTTTTTATTATGTCAGAGCGCGAGCGCCCTCGGCCT 234

Db 181 CTCGCCCATCTCGCTGACTAATTTTTTATTATGTCAGAGCGCGAGCGCCCTCGGCCT 240

QY 235 CTG 237

Db 241 CTG 243

RESULT 5

AX023682

LOCUS AX023682

DEFINITION Sequence 24 from Patent WO0017371.

ACCESSION AX023682

VERSION AX023682.1 GI:10184043

KEYWORDS synthetic construct.

SOURCE synthetic construct

ORGANISM artificial sequences.

REFERENCE 1 (bases 1 to 242)

AUTHORS Binley,K.M. and Naylor,S.

TITLE Polynucleotide constructs and uses thereof

JOURNAL Patent: WO 0017371-A 24 30-MAR-2000;

BINLEY KATIE MARY (GB) ; NAYLOR STUART (GB) ; OXFORD BIOMEDICA LTD (GB)

FEATURES

Location/Qualifiers

1..242

/organism="synthetic construct"

/db_xref="taxon:32630"

/note="Synthetic construct"

BASE COUNT 45 a 93 c 47 g 57 t

ORIGIN

Query Match 73.8%; Score 175; DB 6; Length 242;

Best Local Similarity 84.8%; Pred. No. 2.4e-43;

Matches 196; Conservative 0; Mismatches 35; Indels 0; Gaps 0;

QY 7 GTCGTGACGACGTCGATCTAGTCTGTCGAGGATCTAGTCTGTCGAGCAGTGCAC 66

Db 12 GTCTGCACGACACTAGATGTACGTCCTGTCGACGACACTAGATGTACGTCCTGTCGACGAC 71

QY 67 GTAGCCCGGCTCGAGATCTGGATCTGCATCTCAATTAGTCAGCAACCATAGTCCCG 126

Db 72 TCTAGCCCGGCTCGAGATCTGGATCTGCATCTCAATTAGTCAGCAACCATAGTCCCG 131

QY 127 CCTACTCGGCCCATCCCGCCCTAACTCCGCCAGTTCGCCCATCTCTCGGCCCATC 186

Db 127 CCTACTCGGCCCATCCCGCCCTAACTCCGCCAGTTCGCCCATCTCTCGGCCCATC 191

Db 132

CCCTAACTCCGCCCATCCCGCCCTAACTCCGCCAGTTCGCCCATCTCTCGGCCCATC 191

QY 187

GCTGACTAAATTTTTTTTATTTATTCAGAGCGCGAGCGCGCTCGGCTCTG 237

Db 192

GCTGACTAAATTTTTTTTATTTATTCAGAGCGCGAGCGCGCTCGGCTCTG 242

RESULT 6

RVU57027

LOCUS RVU57027

DEFINITION Reporter vector pCAT<R>3-Promoter vector. linear SYN 24-APR-2001

ACCESSION U57027

VERSION U57027.2 GI:13775608

KEYWORDS Reporter vector pCAT<R>3-Promoter vector.

ORGANISM Reporter vector pCAT<R>3-Promoter vector.

REFERENCE 1 (bases 1 to 4219)

AUTHORS Groskreutz,D.J., Vavra,S., Lesley,S. and Schenborn,E.

TITLE CAT Reporter Systems: New pCAT(R)3 Reporter Vectors and Antibodies

JOURNAL Provide Increased Expression and Detection Capabilities

REFERENCE 2 (bases 1 to 4219)

AUTHORS Promega Notes 55, 2-9 (1996)

TITLE Groskreutz,D.J. and Vavra,S.

JOURNAL Direct Submission

REFERENCE 3 (bases 1 to 4219)

AUTHORS Submitted (30-APR-1996) Production, Promega Corporation, 5445 East

TITLE Cheryl Parkway, Madison, WI 53711, USA

JOURNAL Groskreutz,D.J. and Vavra,S.

REFERENCE 3 (bases 1 to 4219)

AUTHORS Submitted (24-APR-2001) Production, Promega Corporation, 5445 East

TITLE Cheryl Parkway, Madison, WI 53711, USA

REMARK Sequence update by submitter

COMMENT On Apr 24, 2001 this sequence version replaced gi:1399741.

FEATURES

Location/Qualifiers

1..4219

/organism="Reporter vector pCAT<R>3-Promoter vector"

/db_xref="taxon:47943"

misc_feature 1..41

promoter 48..250

misc_feature 185

misc_feature 191

misc_feature 196

intron 295..427

gene 483..1142

CDS 483..1142

/gene="CAT"

/codon_start=1

/transl_table=11

/product="chloramphenicol acetyltransferase"

/db_xref="GI:1399742"

/translation="MEKKITGYTTVDISOWHKEHFEAFQSAQCTYNQVQDITAF

LKTVKKNKHFYPAFIHILARLMAHPELRMAKMDGLVWDSVHPCTYVTHQETTF

SSLMSEYHDDFRQPLHIYSQDVACYGENLAYFPKGTENMEFFYSANPWSTFSDLNV

ANMDFPAPVFTMGKYVTQGDVKVLMPLAIQVHVAVCDFHVGRLNELQQYCDWOGG

A"

1173..1394

/note="SV40 late polyA region"

complement(1452..1481)

/note="RV primer 4 sequencing primer binding site"

1719

/note="ColEI-derived plasmid replication origin"

complement(2481..3341)

/gene="Ampr"

CDS complement(2481..3341)

AUTHORS Kenefick, K.
TITLE Direct Submission
JOURNAL Submitted (05-MAR-2001) Technical Writing, Promega Corporation,
2800 Woods Hollow Road, Madison, WI 53711-5399, USA
REMARK Sequence update by submitter
COMMENT On Mar 5, 2001 this sequence version replaced gi:1200468.
FEATURES
Location/Qualifiers
1..5010
/organism="Cloning vector pGL3-Promoter"
/db_xref="taxon:45860"
/note="luciferase reporter vector"
misc_feature
1..41
/note="multiple cloning site"
48..250
/gene="luc"
280..1932
/gene="luc"
280..1932
/gene="luc"
/codon_start=1
/product="luciferase"
/protein_id="AAA89088.1"
/db_xref="GI:1200469"
/translation="MEDAKNIKKGPAPFYLEDGTAGEQLQHKAMKRYALVPGTIAFTD
AHIEDVITYAEFYFMSGLAEMKRYGLNTNHRIVGSENSLOFFPVLGALRIGYAV
APANDIYNERLLSMGISOPTVVFVSKGLQILNVOKKLPKIIQKIIIMDSKTDVQ
FQSMYFVSHLPGFNEYPFESFDRDKTIALIMNSGSGTGLPKGVLPHERACTVR
FSHARDPIFGNIIIPDAILLSVFPFHGFGMTLGLYGLGCFRWVLMFEELFLES
LDYKIOSALLVPLFSEFAKSTLIDYLSNHEIASGGAPLSKEVEAVAKRFLHP
GIRQYGLTETSAILITPEGDDKPGAVGVFPFEAKVVDLDTGLVGNQGEILCV
RGPYIMSGYVNNPATNALIDKDWLHSGDIAYWDEDEHFFIVDRKLSLIKRYGYVA
PAELIESILLOHPNFDVAGVLPDDAGELPAAVVEHGHKTMTKEIBIVIVASQVTT
AKLRGGVVFVDEYKGLTGKLDARKIREILIKAKGGKIAY"
complement(281..303)
primer_bind
/note="GL primer2 sequencing primer binding site"
1964..2185
polyA_signal
/note="SV40 late poly(A) signal"
complement(2253..2272)
primer_bind
/note="RV primer4 sequencing primer binding site"
2510
rep_origin
/note="ColEI-derived plasmid replication origin"
complement(3272..4132)
gene
/gene="AmpR"
CDS
/gene="AmpR"
/note="ampicillin resistance"
/codon_start=1
/product="beta-lactamase"
/protein_id="AAA89089.1"
/db_xref="GI:1200470"
/translation="MSIQHFRVALIPFFAFLPFAHPETLVKVKDAEDQLGARVGY
IELDLSNGKILESFRPFRPEFPMSTFKVLLCGAVLSRIDAGEQLGRRIRHYSNDLVE
YSPVTEKHLTDGMTVRELCSAATMSDNTAANLLLTITGGPKELTFLHMGDHTVRL
DRWPELNEALPNDREDTTPVMANVTIKLLTGETLLTSLASRQQLIDWMEADKAVGL
LRSLPALPAGWFTADKSGAGERGSRGIIAALGPDGKPSRIVVIYITGSGQATMDERNRQIA
EIGASLIKHW"
4264..4719
/note="fl origin"
rep_origin
4850..5003
polyA_signal
/note="upstream poly(A) signal"
4952..4971
primer_bind
/note="RV primer3 sequencing primer binding site"
BASE COUNT 1288 a 1213 c 1198 g 1311 t
ORIGIN
Query Match 72.2%; Score 171; DB 12; Length 5010;
Best Local Similarity 100.0%; Pred. No. 5.6e-42;
Matches 171; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 67 GCTAGCCCGGGCTCGAGATCTGCATCTCAATTAATAGTCAGCAACACATAGTCCCGC 126
|||||
Db 21 GCTAGCCCGGGCTCGAGATCTGCATCTCAATTAATAGTCAGCAACACATAGTCCCGC 80
|||||
QY 127 CCTAATCCGCCCATCCCGCCCGCCCTAATCCGCCCATCTCCGCCCATCTCCGCCCATC 186
|||||

Db 81 CCTAATCCGCCCATCCCGCCCGCCCTAATCCGCCCATCTCCGCCCATCTCCGCCCATC 140
|||||
QY 187 GCTGACTAATTTTTTTTATTTATGAGAGCGGAGCGCCCTCGGCTCTG 237
|||||
Db 141 GCTGACTAATTTTTTTTATTTATGAGAGCGGAGCGGCTCGGCTCTG 191
|||||

RESULT 11
CVU89938 5115 bp DNA circular SYN 01-AUG-1997
LOCUS Cloning vector pSEAP2-Control, complete sequence.
DEFINITION U89938
ACCESSION U89938
VERSION U89938.1 GI:2190725
KEYWORDS
SOURCE
ORGANISM
REFERENCE
1 (bases 1 to 5115)
unidentified cloning vector.
unidentified cloning vector
artificial sequences: vectors.
AUTHORS
Kitts, P.
TITLE
PSEAP2-Control, complete sequence
JOURNAL
Unpublished
2 (bases 1 to 5115)
Holtz, A. and Lou, Y.
DIRECT SUBMISSION
Submitted (18-FEB-1997) CLONTECH Laboratories, Inc., 1020 East
Meadow Circle, Palo Alto, CA 94303-4230, USA
COMMENT
This vector can be obtained from CLONTECH Laboratories, Inc., 1020
East Meadow Circle, Palo Alto, CA 94303-4230, USA. To place an
order, call (415) 424-8222 or (800) 662-2566, extension 1.
International customers, please contact your local distributor or
subsidiary. For technical information, call (415) 424-8222 or (800)
662-2566, extension 3.
This sequence has been compiled from information in the sequence
databases, published literature and other sources, together with
partial sequences obtained by CLONTECH. If you suspect there is an
error in this sequence, please contact CLONTECH's Technical Support
Department at (415) 424-8222 or (800) 662-2566, extension 3 or
E-mail TECH@CLONTECH.COM.

FEATURES
Location/Qualifiers
1..5115
/organism="unidentified cloning vector"
/db_xref="taxon:45196"
37..245
promoter
/note="SV40 early promoter"
272..1831
CDS
/note="SEAP"
/codon_start=1
/product="secreted alkaline phosphatase"
/protein_id="AAB64402.1"
/db_xref="GI:2190726"
/translation="MLLLLLGLRLQLSLGIIPVEENPDFWNRREAEALGAARKLQ
PAQTAANLIIFLGDGMGYSTVTAARILKQKDKLGPILAMPDFPYVALSKTYNV
DHVPSGATATAYLCGVKNGFTICLSAARFNOCNTRGNEIVSMNRKAKGKSV
GVYTRVHOSAPAGTYATVARNWYSDADVPASARQEGCODIATOLISNMIDIVILG
GGRYRMGTDPPEIPDDYSGGTDLGKLVQEWLAKRQARYWNTETLMQASLD
PSVTHLMGLFPGDMKYEIHRDSTLDPMSLMEATEARLLLSRNPGRFFLFVGGRIDH
GHHSRAYRALTEIMFDARAGLTSEEDTSLVTDADSHSVFSGYPLRGSSIF
GLAPGRKDRKAYTVLLYNGVPGYVLKDGARPDVTESESGSPYEQQSAPVLDDETHA
GEDVAVFARGPQAHVHGVEQFTIAHVMAFAACLEPYTACDLAPAGTTAAHPGYS
RVGAARFSGT"
3377..4237
/gene="AmpR"
complement(3377..4237)
/function="ampicillin resistance"
/codon_start=1
/product="beta-lactamase"
/protein_id="AAB64403.1"
/db_xref="GI:2190727"
/translation="MSIQHFRVALIPFFAFLPFAHPETLVKVKDAEDQLGARVGY
IELDLSNGKILESFRPFRPEFPMSTFKVLLCGAVLSRIDAGEQLGRRIRHYSNDLVE
YSPVTEKHLTDGMTVRELCSAATMSDNTAANLLLTITGGPKELTFLHMGDHTVRL
DRWPELNEALPNDREDTTPVMANVTIKLLTGETLLTSLASRQQLIDWMEADKAVGL
LRSLPALPAGWFTADKSGAGERGSRGIIAALGPDGKPSRIVVIYITGSGQATMDERNRQIA
EIGASLIKHW"

LRSLPAGWFIADKSGAGSRGIIAALPGDKPSRIWYITGSOATMDERNQIA
EIGASLIKHW
4955..5108
/note="transcriptional blocker"

misc_feature

BASE COUNT 1201 a 1392 c 1338 g 1184 t

ORIGIN

Query Match 72.2%; Score 171; DB 12; Length 5115;
Best Local Similarity 100.0%; Pred. No. 5.6e-42;
Matches 171; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 67 GCTAGCCGGCTCGAGATCTCGGATCTGCATCTCAATTAGTCAGCAACCATAGTCCCGC 126
Db 21 GCTAGCCGGCTCGAGATCTCGGATCTGCATCTCAATTAGTCAGCAACCATAGTCCCGC 80
QY 127 CCCTAACTCCGCCATCCCGCCCTAACTCCCGCCAGTCCGCCCATCTCTCGGCCCATC 186
Db 81 CCCTAACTCCGCCATCCCGCCCTAACTCCCGCCAGTCCGCCCATCTCTCGGCCCATC 140
QY 187 GCTGACTAATTTTTTTATTTATGTCAGAGCGCGAGCGCCCTCGGCCCTCTG 237
Db 141 GCTGACTAATTTTTTTATTTATGTCAGAGCGCGAGCGCCCTCGGCCCTCTG 191

RESULT 12

LOCUS AX339208

DEFINITION Sequence 2 from Patent WO0196602.
ACCESSION AX339208
VERSION AX339208.1 GI:18135469

KEYWORDS synthetic construct.

SOURCE synthetic construct.

ORGANISM artificial sequences.

REFERENCE 1

AUTHORS Yang, A. L. and Festing, M.

TITLE Methods and materials to determine the p53 status of a sample by

determining the binding of p53 to a vector

JOURNAL Patent: WO 0196602-A 2 20-DEC-2001;

MEDICAL RESEARCH COUNCIL (GB)

FEATURES Location/Qualifiers

source

1..5256

/organism="synthetic construct"

/db_xref="taxon:32630"

/note="pGL3-Control Vector Sequence"

BASE COUNT 1336 a 1268 c 1281 g 1371 t

ORIGIN

Query Match 72.2%; Score 171; DB 6; Length 5256;
Best Local Similarity 100.0%; Pred. No. 5.6e-42;
Matches 171; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 67 GCTAGCCGGCTCGAGATCTCGGATCTGCATCTCAATTAGTCAGCAACCATAGTCCCGC 126
Db 21 GCTAGCCGGCTCGAGATCTCGGATCTGCATCTCAATTAGTCAGCAACCATAGTCCCGC 80
QY 127 CCCTAACTCCGCCATCCCGCCCTAACTCCCGCCAGTCCGCCCATCTCTCGGCCCATC 186
Db 81 CCCTAACTCCGCCATCCCGCCCTAACTCCCGCCAGTCCGCCCATCTCTCGGCCCATC 140
QY 187 GCTGACTAATTTTTTTATTTATGTCAGAGCGCGAGCGCCCTCGGCCCTCTG 237
Db 141 GCTGACTAATTTTTTTATTTATGTCAGAGCGCGAGCGCCCTCGGCCCTCTG 191

RESULT 13

LOCUS CVU47296

DEFINITION Cloning vector pGL3-Control, DNA linear SYN 17-APR-2002

ACCESSION U47296

VERSION U47296.2 GI:13195704

KEYWORDS Cloning vector pGL3-Control.

SOURCE

ORGANISM

Cloning vector pGL3-Control

artificial sequences; vectors.

REFERENCE 1 (bases 1 to 5256)

AUTHORS Groskreutz, D.J. and Schenborn, E.T.

TITLE Direct Submission

JOURNAL Submitted (26-JAN-1996) D.J. Groskreutz, R&D, Promega Corporation,

5445 East Cheryl Parkway, Madison, WI 53711, USA

REFERENCE 2 (bases 1 to 5256)

AUTHORS Kenefick, K.

TITLE Direct Submission

JOURNAL Submitted (05-MAR-2001) Technical Writing, Promega Corporation,

2800 Woods Hollow Road, Madison, WI 53711-5399, USA

REMARK Sequence update by submitter

COMMENT On Mar 5, 2001 this sequence version replaced gi:1200462.

FEATURES Location/Qualifiers

source

1..5256

/organism="Cloning vector pGL3-Control"

/db_xref="taxon:45858"

/note="luciferase reporter vector"

misc_feature

1..41

/note="multiple cloning site"

promoter

48..250

/product="luciferase"

gene

280..1932

/gene="luc"

CDS

280..1932

/gene="luc"

/codon_start=1

/product="luciferase"

/protein_id="AAA89084.1"

/db_xref="GI:1200463"

/translation="MEDAKNIKGPAPFPLEDTAGBQLHKAMRYALVPQTIAFTD

AHEVDITYAEFEMSVRLAEAMKRYGLNTHRIVVCNSLQTFMPLVGLFVAV

AFANDIYNRELLNSMGTSQPTVVFVSKGLQKILNVQKLPKIIKIMDSKTDYQG

FQSMYTFVTHLPFGFNEFDYFSPDRDKTIALIMNSGSLGKGLVAPLHRTACVR

FSHARDPIFGNIIPTDALLSVVPHFGHGMFTTGLYLCGFRVLMVRFEREFLRS

LDYKIQSALLVPTLFSFPAKSTLDIKVDLSNLHETASGAPLSKEVEANAKRHLP

GIQGVGLTETTSAILIPEGDKGKGVVPPFEAKVVDLDTGKTLGVNKGELCV

RGPIMSGYVNPENATNALIDKDGWLHSDIAYWDEDEHFFIVDLKSLIKYKQOVA

PAKLESLIQHNFIDGAVGLPDDAGELPAAVVLEHGKTMTEKEIVDYVASQVTT

AKLRGGVVFVDEVPKGLTGKLDARKIREILIKAKKGGKIAV"

complement(281..303)

/note="GL primer2 sequencing primer binding site"

primer_bind

1964..2185

/note="SV40 late poly(A) signal"

polyA_signal

2205..2441

complement(2499..2518)

enhancer

primer_bind

2756

/note="RV primer4 sequencing primer binding site"

rep_origin

complement(3518..4378)

gene

/note="ColEI-derived plasmid replication origin"

complement(3518..4378)

CDS

complement(3518..4378)

/gene="AmpR"

/note="ampicillin resistance"

/codon_start=1

/product="beta-lactamase"

/protein_id="AAA89085.1"

/db_xref="GI:1200464"

/translation="MSIOHFRVALIPFAAFCLPVFAHPETLVKVKDAEDQLGARVGY

IELDLSNGKILESFEPRFFPMSTFKVLLCGAVLSRIDAGQEQIGRRIRHYSONDLVY

YSPVTEKHLTDGTVVRELCSAATMSDNTAANLLTTIGGPKELTAFLLHNGDHVRL

DRWPELNEATPNDRDTPVAVATTLRLKLLTGELTLASRQQLIDWMEADKRVAGPL

LRSLPAGWFIADKSGAGSRGIIAALPGDKPSRIWYITGSOATMDERNQIA

EIGASLIKHW"

4510..4965

/note="f1 origin"

5096..5249

/note="upstream poly(A) signal"

5198..5217

/note="RV primer3 sequencing primer binding site"

BASE COUNT 1337 a 1268 c 1281 g 1370 t

ORIGIN

```
Query Match 72.2%; Score 171; DB 12; Length 5256;
Best Local Similarity 100.0%; Pred. No. 5.6e-42;
Matches 171; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 67 GCTAGCCGGCTCGAGATCTCGATCTGCATCTCAATTAGTCAGCAACCATAGTCCCGC 126
DB 21 GCTAGCCGGCTCGAGATCTCGATCTGCATCTCAATTAGTCAGCAACCATAGTCCCGC 80
QY 127 CCCTAACTCCGCCATCCCGCCCTAACTCCGCCAGTTCGCCGCCATCTCTCGGCCCATC 186
DB 81 CCCTAACTCCGCCATCCCGCCCTAACTCCGCCAGTTCGCCGCCATCTCTCGGCCCATC 140
QY 187 GCTGACTAATTTTTTTTATTTATGACAGAGCCGAGCGCCCTCGGCTCTG 237
DB 141 GCTGACTAATTTTTTTTATTTATGACAGAGCCGAGCGCCCTCGGCTCTG 191

RESULT 14
AB038600 6320 bp DNA circular SYN 22-FEB-2001
LOCUS Cloning vector pVLUC441 DNA, complete sequence.
DEFINITION AB038600
ACCESSION AB038600
VERSION AB038600.1 GI:13094136
KEYWORDS beta-lactamase; luciferase.
SOURCE Cloning vector pVLUC441 DNA.
ORGANISM Cloning vector pVLUC441
artificial sequences: vectors.
REFERENCE 1 (sites)
AUTHORS Hashinaka,K.
TITLE Synthetic Autonomous Vectors Based on Palindromic Sequences of Parvovirus B19
JOURNAL Published Only in DataBase (2001)
REFERENCE 2 (bases 1 to 6320)
AUTHORS Hashinaka,K.
TITLE Direct Submission
JOURNAL Submitted (21-FEB-2000) Kazuya Hashinaka, Miyazaki Medical College, Department of Biochemistry; 5200 Kihara, Kiyotake, Miyazaki 889-1692, Japan (E-mail:hashinaka@postl.miyazaki-med.ac.jp, Tel:81-985-85-0985, Fax:81-985-85-2401)

FEATURES
source 1. .6320
/organism="Cloning vector pVLUC441"
/db_xref="taxon:117918"
/focus
source 6. .479
/organism="B19 virus"
/notes="taxon:10798"
/db_xref="taxon:Parvovirus B19"
source 2937. .3420
/organism="B19 virus"
/notes="taxon:10798"
/db_xref="taxon:Parvovirus B19"
repeat_region 9. .391
gene 1347. .2414
/genes="luc"
CDS 1347. .2414
/genes="luc"
/transl_table=11
/product="luciferase"
/protein_id="BAB32737.1"
/db_xref="GI:13094137"
/translation="MNSGSGTGLPKGVLPHTACVRFSHARDPIFGNQIIPDTAILS
VVFPHFGMFTTGLICGFRVLMYREELFLRSLQDYKIQSALIVPTLFSFPAK
STLIDYDLSNLHAIASGGAPLSKEVEAVAKRHLPGTROYGLTETTSAITLPEG
DDPAGVKVPPFEAKVDLDTGKTLGYNQRCGLCVRGPMIMSGYVNNPEATNALID
KDWLHSGDIAVDEDEHEFFIVDLRLSLIKYKIQVAPAELESILLQHPNIFDAGVAC
LPDDAGELPAAVVDVLEHGKTMTEKEIVDIVASQVTTAKKRGGVVVDVDEPKGLTGK
LDARKIREILIKAKGCKTAV"
repeat_region 3029. .3411
gene 4252. .5112
/genes="ampR"
CDS 4252. .5112
```

```
BASE COUNT 1580 a 1527 c 1589 g 1624 t
ORIGIN
Query Match 67.5%; Score 160; DB 12; Length 6320;
Best Local Similarity 99.4%; Pred. No. 1.4e-38;
Matches 171; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

QY 67 GCTAGCCGGCTCGAGATCTCGATCTGCATCTCAATTAGTCAGCAACCATAGTCCCGC 126
DB 495 GCTAGCCGGCTCGAGATCTCGATCTGCATCTCAATTAGTCAGCAACCATAGTCCCGC 554
QY 127 CCCTAACTCCGCCATCCCG--CCCTAACTCCGCCAGTTCGCCGCCATCTCTCGGCCCAT 185
DB 555 CCCTAACTCCGCCATCCCGACCCTAACTCCGCCAGTTCGCCGCCATCTCTCGGCCCAT 614
QY 186 GCCTGACTAATTTTTTTTATTTATGACAGAGCCGAGCGCCCTCGGCTCTG 237
DB 615 GCCTGACTAATTTTTTTTATTTATGACAGAGCCGAGCGCCCTCGGCTCTG 666

RESULT 15
AB037684 5917 bp DNA circular SYN 04-APR-2000
LOCUS Cloning vector pPLuc, complete sequence.
DEFINITION AB037684
ACCESSION AB037684
VERSION AB037684.1 GI:7415874
KEYWORDS Pluc; luciferase; modified C.
SOURCE synthetic construct DNA.
ORGANISM synthetic construct
artificial sequences.
REFERENCE 1 (bases 1 to 5917)
AUTHORS Chen,A.B., Kao,A.Y.-F. and Brown,C.M.
TITLE A short open reading frame within the encapsidation signal affects the translation of the polymerase gene from hepatitis Bvirus
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 5917)
AUTHORS Chen,A.B., Kao,A.Y.-F. and Brown,C.M.
TITLE Direct Submission
JOURNAL Submitted (27-JAN-2000) Chris M. Brown, University of Otago, Department of Biochemistry; P.O.Box 56, Dunedin, Otago 9001, New Zealand (E-mail:chris.brown@stonebow.otago.ac.nz, Tel:+64-3-479-7875, Fax:+64-3-479-7866)

FEATURES
source 1. .5917
/organism="synthetic construct"
/db_xref="taxon:32630"
/notes="cloning vector pPLuc"
misc_feature 1. .495
/notes="HBV Adw 5' leader of pregenomic RNA"
CDS 35. .94
/notes="ORF"
/organism="synthetic construct"
/translation="MNSGSGTGLPKGVLPHTACVRFSHARDPIFGNQIIPDTAILS
VVFPHFGMFTTGLICGFRVLMYREELFLRSLQDYKIQSALIVPTLFSFPAK
STLIDYDLSNLHAIASGGAPLSKEVEAVAKRHLPGTROYGLTETTSAITLPEG
DDPAGVKVPPFEAKVDLDTGKTLGYNQRCGLCVRGPMIMSGYVNNPEATNALID
KDWLHSGDIAVDEDEHEFFIVDLRLSLIKYKIQVAPAELESILLQHPNIFDAGVAC
LPDDAGELPAAVVDVLEHGKTMTEKEIVDIVASQVTTAKKRGGVVVDVDEPKGLTGK
LDARKIREILIKAKGCKTAV"
repeat_region 3029. .3411
gene 4252. .5112
/genes="ampR"
CDS 4252. .5112
```


GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 3, 2003, 22:54:17 ; Search time 3151.28 Seconds
(without alignments)
1218.024 Million cell updates/sec

Title: US-09-787-562-9

Perfect score: 237

Sequence: I gctagatctgtcaggacgt.....cgaggcgcctgcgcctctg 237

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

```

1: em_estba:*
2: em_esthum:*
3: em_estlin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_esttc:*
9: gb_estti:*
10: gb_estt2:*
11: gb_htc:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: gb_gss:*
18: em_gss_hum:*
19: em_gss_inv:*
20: em_gss_pln:*
21: em_gss_vrt:*
22: em_gss_fun:*
23: em_gss_mam:*
24: em_gss_mus:*
25: em_gss_other:*
26: em_gss_pro:*
27: em_gss_rod:*

```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	ID	Description
C 1	80.8	34.1	569	12	BG063359
C 2	79.8	33.7	392	9	AA078277
C 3	55	23.2	578	9	AA409312
C 4	41	17.3	869	10	BE216820
C 5	38.2	16.1	974	17	A2157908
C 6	37.6	15.9	868	17	CNS03VAE

C	7	37.2	15.7	684	12	BG620549
	8	37.2	15.7	725	17	AG062985
C	9	37	15.6	639	17	B19583
	10	37	15.6	862	17	AG137051
	11	37	15.6	1619	14	BM906235
	12	36.8	15.5	371	13	B1885730
	13	36.8	15.5	911	17	A2682313
C	14	36.4	15.4	522	17	CNS015B2
C	15	36.4	15.4	978	17	CNS018FK
	16	36.2	15.3	453	13	B1885726
	17	36.2	15.3	518	17	A2079405
	18	36.2	15.3	830	17	B19180
	19	36.2	15.3	1101	17	CNS0181N
	20	36	15.2	552	17	B19569
C	21	36	15.2	629	10	BE239175
C	22	36	15.2	631	10	AW727988
	23	36	15.2	982	17	CNS014D5
	24	36	15.2	1137	13	BM415579
	25	35.6	15.0	755	17	AG100419
C	26	35.6	15.0	1087	17	AG147408
	27	35.6	15.0	1101	17	CNS0036H
	28	35.4	14.9	309	13	B1885645
	29	35.4	14.9	866	12	BG442104
	30	35.2	14.9	625	17	B19473
	31	35.2	14.9	870	17	B157873
	32	35.2	14.9	892	17	BH133618
C	33	35.2	14.9	929	17	A2687353
	34	35.2	14.9	971	17	BH156534
	35	35.2	14.9	1093	13	BM544531
	36	35	14.8	398	12	BG364630
	37	35	14.8	449	13	B1885612
C	38	35	14.8	575	9	AA800583
	39	35	14.8	802	17	AG139568
	40	35	14.8	846	17	B18514
	41	35	14.8	954	13	BF833711
	42	35	14.8	1192	17	B11834
	43	35	14.8	1211	14	BM807956
	44	35	14.8	1389	14	BM924250
	45	34.8	14.7	657	17	AG111841

ALIGNMENTS

RESULT 1
BG063359/c
LOCUS
DEFINITION
H3006C07-3 NTA Mouse 15K cDNA Clone Set Mus musculus cDNA clone
H3006C07 3', mRNA sequence.
ACCESSION
BG063359
KEYWORDS
EST.
SOURCE
house mouse.
ORGANISM
Mus musculus
REFERENCE
1 (bases 1 to 569)
AUTHORS
Kargul, G.J., Dudekula, D.B., Qian, Y., Lim, M.K., Jaradat, S.A., Tanaka, T.S., Carter, M.G. and Ko, M.S.H.
TITLE
Verification and initial annotation of NIA mouse 15K cDNA clone set
JOURNAL
Unpublished (2001)
COMMENT
Contact: George J. Kargul
Laboratory of Genetics
National Institute on Aging/National Institutes of Health
333 Cassell Drive, Suite 4000, Baltimore, MD 21224-6820, USA
Email: cdnae@igsun.grc.nia.nih.gov
This clone set has been freely distributed to the community. Please visit <http://igsun.grc.nia.nih.gov/cDNA/15K.html> for details.
Plate: H3006 row: C column: 07
Seq primer: -21M13 Forward
High quality sequence stop: 569
POLYA-No.

FEATURES

Location/Qualifiers

BG620549 602619705
AG062985 Pan trogl
B19583 F1114-T7 IG
AG137051 Pan trogl
BM906235 AGENCOURT
B1885730 fn12e08.Y
A2682313 ENTKQ34TF
AL105257 Drosophil
AL109274 Drosophil
B1885726 fn12d03.Y
A2079405 RPT-23-4
B19180 F3K9-T7 IG
AL108773 Drosophil
B19569 F7A20-T7.1
BE239175 MD0831 Me
AW727988 GA_Ea002
AL104003 Drosophil
BM415579 OP20656 M
AG100419 Pan trogl
AG147408 Pan trogl
AL064420 Drosophil
B1885645 fn09g02.Y
BG442104 GA_Ea001
B19473 T5L16-T7 TA
B157873 ENTRB41TR
BH133618 ENTVOV1TR
A2687353 ENTJ25TF
BH156534 ENTSE18TR
BM544531 AGENCOURT
BG364630 dc74a08.Y
B1885612 fn08e02.Y
AA800583 EST190080
AG139568 Pan trogl
B18514 F20L4-T7 IG
B1833711 603088435
B11834 F16E8-SP6 I
BM807956 AGENCOURT
BM924250 AGENCOURT
AG111841 Pan trogl

uninoculated leaves were harvested 20 hr post-inoculation (Wei, Wise). In the TJ Close lab at the University of California, Riverside, total RNA was prepared from each sample pool, equal quantities of all three RNA pools were combined, poly(A) RNA was purified from the mixture, one pfu were in vivo excised to give pBluescript SK(-) cDNA phagemids (Choi, Close). Phagemids were plated and picked at the Clemson University Genomics Institute (CUGI) (Begum , Palmer, Frisch, Atkins and Wing). Plasmid DNA preparations, DNA sequencing and sequence analysis were performed at CUGI (Wing, Yu, Frisch, Henry, Simmons, Oates , Rambo, Main). The sequence has been trimmed to remove vector sequence and contains a minimum of 100 bases of phred value 20 or above. For more details on library preparation and sequence analysis see <http://www.genome.clemson.edu/projects/barley>. To order this clone see <http://www.genome.clemson.edu/orders> Also see Close TJ, Wing R, Kleinhofs A, Wise R (2001) Genetically and physically anchored EST resources for barley genomics. Barley Genetics Newsletter 31:29-30. (<http://wheat.pw.usda.gov/ggpages/bgn/31/cover.html>)."

BASE COUNT 174 a 213 c 367 g 115 t
ORIGIN

Query Match 17.3%; Score 41; DB 10; Length 869;
Best Local Similarity 58.7%; Pred. No. 3.3;
Matches 71; Conservative 0; Mismatches 50; Indels 0; Gaps 0;

QY 115 CCATAGTCGGCCGCTAACTCCGCCCATCGCCGCCTTACTCGCCGACAGTTCGGCCATT 174
 || | ||| |||| | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 837 CCTTCGCCGCCGCCCTACCCTCCGCTCCGCCCTCCGCCCTTTCCCTCCTCT 778

QY 175 CTCGCGCCCATCGCTGACTCAATTTTTTTTATTTATCGAGAGCGGCGGCGCTCGGCCT 234
 ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 777 CTCTCGCTCCCGCACCTACCCCTTATTACTTTTCACCGCGCCCTCGCCACCTCCC 718

QY 235 c 235
 |
Db 717 c 717

RESULT 5
AZ157908

LOCUS AZ157908 linear GSS 29-AUG-2000
DEFINITION SP_0059_A1_D06_T7A Strongylocentrotus purpuratus, purple sea urchin
 sperm genomic BAC library Strongylocentrotus purpuratus genomic
 clone plate=59 Col=11 Row=G, DNA sequence.

ACCESSION AZ157908
VERSION AZ157908.1 GI:8310511
KEYWORDS GSS.
SOURCE Strongylocentrotus purpuratus.
ORGANISM Strongylocentrotus purpuratus
Eukaryota; Metazoa; Echinodermata; Eleutherozoa; Echinozoa;
Echinoidea; Euechinoidea; Echinacea; Echinoidae;
Strongylocentrotidae; Strongylocentrotus.
l (bases 1 to 974)
Cameron,R.A., Maharas,G., Rast,J.P., Martinez,P., Biondi,T.R.,
Swartzell,S., Wallace,J.C., Poustka,A.J., Livingston,B.T., Wray,
G.A., Ettensohn,C.A., Lehrach,H., Britten,R.J., Davidson,E.H. and
Hood,L.

TITLE A sea urchin genome project: Sequence scan, virtual map, and
additional resources
Proc. Natl. Acad. Sci. U.S.A. 97 (17), 9514-9518 (2000)
20402566
Contact: Cameron, RA, Davidson, EH, Hood, L
Division of Biology 156-29
California Institute of Technology
Pasadena California 91125, USA
Tel: (626) 395-8421
Fax: (626) 793-3047
Email: acameron@caltech.edu
Plate: 59 row: G column: 11

QY 72 CCGGGCTCGAGATCTGGATCTGCATCTCAATTAGTCAGCAACCATAGTCCGGCCCTA 131
DB 550 CCGGCTCTTGTCTGGGGGCTGTCTTTTCTCGGTTCCTCCCGCCCTCCCTCGGCTC 491
QY 132 ATTCGGCCGATCCGGGCCCTTAACCTCGGCCCATGTTCCGCCCATTCCTCGGCCCATCGCTGA 191

```

FEATURES
  source
    Location/Qualifiers
      1..868
        /organism="tetraodon nigroviridis"
        /db_xref="taxon:99883"
        /clone="061P09"
        /clone_lib="G"
        /note="Genoscope sequence ID : C03G061CH05spl-end :
        pUC-ori"

```



```

Db 490 GCGGCCCCCCCCCGGCTTCGGCGCGCTTCGGCGGCCCCCGCCCCCGCGCTCC 431
QY 192 CTAATTTTTTTTATTTAT 209
Db 430 TGCTTTTTTTCTTTCT 413

RESULT 8
AG062985
LOCUS Pan troglodytes DNA, clone: PTB-051101.F, genomic survey sequence.
DEFINITION AG062985
ACCESSION AG062985
VERSION AG062985.1 GI:16614787
KEYWORDS GSS.
SOURCE Pan troglodytes male lymphoblast DNA, clone_lib:PTB Chimpanzee Male
BAC Library clone:PTB-051101.F.
ORGANISM
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.
REFERENCE
AUTHORS Fujiyama, A., Hattori, M., Toyoda, A., Taylor, T.D., Yada, T.,
Totoki, Y., Watanabe, H. and Sakaki, Y.
TITLE BAC end sequences of Library PTB
JOURNAL Unpublished
REFERENCE
AUTHORS 2 (bases 1 to 725)
Fujiyama, A., Hattori, M., Toyoda, A., Taylor, T.D., Yada, T.,
Totoki, Y., Watanabe, H. and Sakaki, Y.
TITLE Direct Submission
JOURNAL Submitted (02-AUG-2001) Asao Fujiyama, The Institute of Physical
and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
1-7-22 Suehiro-chou, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
(E-mail: chimbes@gsc.riken.go.jp, URL: http://hgp.gsc.riken.go.jp/,
Tel: 81-45-503-9111, Fax: 81-45-503-9170)
COMMENT Clones are derived from the chimpanzee BAC library PTB This BAC end
was generated during the R&D process and may have higher chance of
clone tracking errors.
PRIMERS
Sequencing: -21M13
LIBRARY
Vector : pKSL45
R.Site 1 : SacI
R.Site 2 : SacI.
FEATURES
source
Location/Qualifiers
1..725
/organism="Pan troglodytes"
/db_xref="taxon:9598"
/clone="PTB-051101.F"
/sex="male"
/cell_type="lymphoblast"
/clone_lib="PTB Chimpanzee Male BAC Library"
BASE COUNT 55 a 423 c 51 g 195 t 1 others
ORIGIN

Query Match 15.7%; Score 37.2; DB 17; Length 725;
Best Local Similarity 57.9%; Pred. No. 28;
Matches 66; Conservative 0; Mismatches 48; Indels 0; Gaps 0;

QY 92 TCTGCATCTCAATTAGTCAGCAACCATAGTCCGCGCCCTTAACCTCCGCCATCCCGCCCT 151
Db 550 TCCTTCTCCCTCTTCCCTCCCGCCCTCTCTCCCTCCCGCCCTCTCCCTCCCGCCCTCAT 609
QY 152 AACCTCCGCCAGTTCGCGCCATCTCCGCGCCCATCGCTGACTAATTTTTTTAT 205
Db 610 TCCTTTTCCCGCTCCCGCTCTCTCCCGCCCGCCCTCTCTCTTATTTCTTTCT 663

RESULT 9
B19583
LOCUS B19583
DEFINITION FI114-T7 IGF Arabidopsis thaliana genomic clone FI114, DNA
sequence.
ACCESSION B19583

```

```

B19583.1 GI:2394637
GSS.
thale cress.
Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
1 (bases 1 to 639)
Feng, J., Dewar, K., Buehler, E., Kim, C., Li, Y., Shinn, P., Sun, H. and
Ecker, J.
BAC End Sequences at ATGC
Unpublished (1997)
Other GSSs: FI114-Sp6.1, FI114-Sp6, FI114-Sp6.2, FI114-Sp6.3,
FI114-Sp6.4
Contact: Ecker J.
Arabidopsis Thaliana Genome Center
University of Pennsylvania
Dept. of Biology, University of Pennsylvania, Philadelphia, PA
19104
Tel: 215-898-9384
Fax: 215-898-8780
Email: Jecker@atgenome.bio.upenn.edu
Seq primer: T7
Class: BAC ends
High quality sequence start: 458
High quality sequence stop: 476.
Location/Qualifiers
1..639
/organism="Arabidopsis thaliana"
/strain="Columbia"
/db_xref="taxon:3702"
/clone="FI114"
/clone_lib="IGF"
/sex="hermaphrodite"
/note="Vector: B19583; Site_1: EcoRI; Site_2: EcoRI;
Produced by Thomas Altmann"
BASE COUNT 81 a 365 c 32 g 130 t 31 others
ORIGIN

Query Match 15.6%; Score 37; DB 17; Length 639;
Best Local Similarity 54.7%; Pred. No. 32;
Matches 70; Conservative 0; Mismatches 58; Indels 0; Gaps 0;

QY 108 TCAGCAACCATAGTCCGCGCCCTAAGTCCGCCATCCCGCCCTAAGTCCGCCAGTTCC 167
Db 373 TCCTCCCGCTCCCGCCCTCCCTTCCTCCCGCCCGCCCGCCCTCCCGCTTC 432
QY 168 GCCCATCTCCGCGCCCATCGCTGACTAATTTTTTTATTTATGACAGAGCCGAGCGCC 227
Db 433 NCCTCCCGCCCGCCCGCCCTTTTTTTTTTTTTTTTTTTTNCCTCCCGCCCGCC 492
QY 228 TCGGCTC 235
Db 493 CCNCCGCC 500

RESULT 10
AG137051/c
LOCUS AG137051
DEFINITION Pan troglodytes DNA, clone: PTB-150N02.F, genomic survey sequence.
ACCESSION AG137051
VERSION AG137051.1 GI:16666729
KEYWORDS GSS.
SOURCE Pan troglodytes male lymphoblast DNA, clone_lib:PTB Chimpanzee Male
BAC Library clone:PTB-150N02.F.
ORGANISM
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.
REFERENCE
AUTHORS Fujiyama, A., Hattori, M., Toyoda, A., Taylor, T.D., Yada, T.,
Totoki, Y., Watanabe, H. and Sakaki, Y.
TITLE BAC end sequences of Library PTB
JOURNAL Unpublished

```



```

/organism="Danio rerio"
/db_xref="taxon:7955"
/clone_lib="zebrafish Washu MPIMG EST"
/sex="mixed"
/tissue_type="26 somite embryos, adult livers, shield
stage embryos"
/lab_host="X11-blue MRF"
/notes="Vector: pSPORT1; Site_1: NotI; Site_2: SalI; 1st
strand cDNA was primed with a Not I - oligo(dT)15 primer
[5'-GCAGTACCTCTAGATCGGAGCGCCGCTTTTCTTTT3'];
double-stranded cDNA was ligated to Sal I adaptors (BRL),
digested with Not I and cloned into the Not I and Sal I
sites of the pSPORT1 vector (BRL). Library was constructed
by Matthew Clark (Lehrach lab; ICRP, London and Max Planck
Institut fuer Molekulare Genetik, Berlin). cDNAs for EST
analysis were selected following oligonucleotide
hybridization fingerprinting of arrayed clones from
zebrafish late somitogenesis (26 ss), adult liver or
embryonic shield stage (5.6 h) libraries. Fingerprint
data were used to computationally cluster cDNAs, and a
single cDNA from each cluster was chosen for sequencing.
In some cases multiple members of the same cluster were
sequenced to assess clustering parameters or single clones
were sequenced additional times to assess quality
control."
BASE COUNT      5 a 298 c 6 g 37 t 25 others
ORIGIN

Query Match      15.5%; Score 36.8; DB 13; Length 371;
Best Local Similarity 57.0%; Pred. NO. 34;
Matches 65; Conservative 0; Mismatches 49; Indels 0; Gaps 0;

QY 122 CCGCGCCCTAATCGCCGCTTCTTATTTATTCAGAGCGCGCTCGGCTC 181
    ||| ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 128 CCGCGCCCTTCTCGCCGCTTCTTATTTATTCAGAGCGCGCTCGGCTC 187
    ||| ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 182 CCATCGCTGACTAATTTTATTTATTCAGAGCGCGCTCGGCTC 235
    || | | | | ||||| ||| | | | ||||| ||| ||| ||| ||| |||
Db 188 CCGCGCCGCTTCTTCTTATTTTCTTCTTCTTCTTCTTCTTCTTCTT 241
    || | | | | ||||| ||| | | | ||||| ||| ||| ||| ||| |||

RESULT 13
A2682313
LOCUS
DEFINITION
  ENFKQ34TF Entamoeba histolytica Sheared DNA Entamoeba histolytica
  genomic, DNA sequence.
ACCESSION
  A2682313
KEYWORDS
  GSS.
SOURCE
  Entamoeba histolytica.
  Entamoebidae; Entamoeba.
REFERENCE
  1 (bases 1 to 911)
AUTHORS
  Loftus,B., Van Aken,S. and Fraser,C.
TITLE
  Determination of clone end sequences from Entamoeba histolytica
  HMI:IMSS sheared DNA library
JOURNAL
  Unpublished (2000)
COMMENT
  Contact: Brendan J Loftus
  Department of Eukaryotic Genomics
  The Institute for Genomic Research
  9712 Medical Center Dr., Rockville, MD 20850, USA
  Tel: 301 838 0208
  Fax: 301 838 3543
  Email: bjlloftus@tigr.org
  Clones are derived from the Entamoeba histolytica HMI:IMSS sheared
  DNA library
  Seq primer: M13-Forward
  Class: shotgun
  High quality sequence start: 30
  High quality sequence stop: 330.
  Location/Qualifiers
    1..911

```

```

/organism="Entamoeba histolytica"
/strain="HMI:IMSS"
/db_xref="taxon:5759"
/clone_lib="Entamoeba histolytica Sheared DNA"
/notes="Vector: pSPORT1; Site_1: Bst I; Constructed at The
Institute for Genomic Research (TIGR), Rockville, MD.
Genomic DNA isolated from broth cultures of E. histolytica
using a method described by Clark and Diamond (Clark,
C.G., and Diamond, L.S. (1993) Entamoeba histolytica: a
method for isolate identification. Exp. Parasitol.
77:450.). The DNA was mechanically sheared to give a
tight size distribution (~2 kb). The v + i method used for
the library construction is described in detail in Smith,
H.O. and Venter, J.C. (Making small insert libraries for
whole genome shotgun sequencing projects. In Genome
Sequencing: A Practical Approach, eds. M. Vaudin and B.
Barell, Oxford University Press, 1999)."
BASE COUNT      171 a 117 c 74 g 549 t
ORIGIN

Query Match      15.5%; Score 36.8; DB 17; Length 911;
Best Local Similarity 63.6%; Pred. No. 36;
Matches 56; Conservative 0; Mismatches 32; Indels 0; Gaps 0;

QY 122 CCGCGCCCTAATCGCCGCTTCTTATTTATTCAGAGCGCGCTCGGCTC 181
    ||| ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 373 CCGCGCCGCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTT 432
    ||| ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 182 CCATCGCTGACTAATTTTATTTATTTATTTATTTATTTATTTATTTAT 209
    || | | | | ||||| ||| | | | ||||| ||| ||| ||| ||| |||
Db 433 CCGCGCCGCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTT 460
    || | | | | ||||| ||| | | | ||||| ||| ||| ||| ||| |||

RESULT 14
CNS015BZ/c
LOCUS
DEFINITION
  Drosophila melanogaster genome survey sequence sp6 end of BAC
  BACN13C14 of DrosBAC library from Drosophila melanogaster (fruit
  fly), genomic survey sequence.
ACCESSION
  AL105257
VERSION
  AL105257.1 GI:5617271
KEYWORDS
  GSS.
SOURCE
  Drosophila melanogaster.
  Drosophila melanogaster.
  Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
  Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
  Ephydroidea; Drosophilidae; Drosophila.
REFERENCE
  1 (bases 1 to 522)
AUTHORS
  Genoscope.
TITLE
  Direct Submission
JOURNAL
  Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage :
  BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
  - Web : www.genoscope.cns.fr)
COMMENT
  Determination of this BAC-end sequence was carried out as part of a
  collaboration with the European Drosophila Genome Project (EDGP) -
  http://www.edgp.ebi.ac.uk. This Drosophila melanogaster BAC
  library (Dros BAC) was made by Alain Billaud at CPH (Centre
  d'Etude du Polymorphisme Humain) with funding provided by a MRC
  project grant. The DNA was prepared from embryos by Alain Bucheton
  and Genevieve Payan. It has been constructed in the vector
  pBelobAC11.
FEATURES
  Location/Qualifiers
    1..522
  /organism="Drosophila melanogaster"
  /db_xref="taxon:7227"
  /clone_lib="BACN13C14"
  /clone_lib="DrosBAC"
  /plasmid="pBelobAC11"
  /note="end : SP6"
BASE COUNT      89 a 32 c 303 g 47 t 51 others
ORIGIN

Query Match      15.4%; Score 36.4; DB 17; Length 522;

```

Search completed: January 4, 2003, 01:04:18
Job time : 3158.28 secs

[illegible]

	1	2	3	4	5	6	7	8	9	
	237	100.0	237	21	AAAL2001	Murine PGK HRE der				
	221	93.2	243	20	AAAL1397	Murine PGK fragmen				
	221	93.2	243	21	AAAL1985	Murine PGK HRE der				
	209	88.2	243	20	AAAL2078	Promoter Obhrel us				
	175	73.8	242	21	AAAL2016	Murine PGK HRE der				
	171	72.3	5010	24	AAAL2753	gGL3 promoter vect				
	171	72.2	5256	21	AAAL07776	DNA sequence of pl				
	171	72.2	5256	24	AAAL27537	gGL3 control vecto				
	152.2	64.2	267	21	AAAL2019	Human erythropoiet				

```
XX Novel polynucleotide constructs comprising at least two repeats of a
PT hypoxia response element useful for driving expression of nucleic acids
PT of interest in a cell under hypoxic conditions
XX
XX Example 1; Page 68; 155pp; English.
XX
XX This invention describes novel polynucleotide comprising at least 2
CC repeats of a hypoxia response element (HRE), where the hypoxia-inducible
CC factor (HIF) consensus binding sites within each of the 2 repeats are
CC separated by a spacer of at least 20 contiguous nucleotides. The products
CC of the invention have vasotropic, cardiatic, cytostatic and antiarthritic
CC activity and can be used for gene therapy. The polynucleotides are useful
CC for delivering nucleic acids of interest to mammalian cells. Lentiviral
CC vectors are responsive to hypoxic agents and to agents that mimic
CC hypoxia. This regulation can be harnessed in vitro to enhance the
CC production of the vector and can be used in vivo to regulate gene
CC expression in response to a physiological signal. The vectors have
CC utility in disease, where ischaemia, including hypoxia, is a feature,
CC e.g. cardiovascular disease, peripheral arterial disease, cancer and
CC arthritis. The novel regulatory construct is capable of driving very high
CC levels of transcription under conditions of hypoxia whilst providing only
CC low basal levels of transcription under normal oxygen conditions. The
CC polynucleotide construct targets cells within a tumor mass that are under
CC conditions of hypoxia without affecting normal surrounding tissue. This
CC sequence represents a murine phosphoglycerate kinase (PGK) HRE derived
CC promoter Obhrell which is described in the method of the invention.
XX
XX Sequence 237 BP; 43 A; 82 C; 56 G; 56 T; 0 other;
SQ
Query Match 100.0%; Score 237; DB 21; Length 237;
Best Local Similarity 100.0%; Pred. No. 6.3e-63;
Matches 237; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GCTAGAGTCTGTCAGGACGTGACATCTAGTGTGTCGAGGATCTAGTGTGTCGAGGAC 60
Db 1 GCTAGAGTCTGTCAGGACGTGACATCTAGTGTGTCGAGGATCTAGTGTGTCGAGGAC 60
QY 61 GTGACAGTACGCGGCTGAGATCTGATCTGATCTCAATTAGTCAGCAACCATAG 120
Db 61 GTGACAGTACGCGGCTGAGATCTGATCTGATCTCAATTAGTCAGCAACCATAG 120
QY 121 TCCGCGCCCTAACTCCGCCCATCCGCCCTAACTCCGCCCATCTCCGCCCATCTCCGC 180
Db 121 TCCGCGCCCTAACTCCGCCCATCCGCCCTAACTCCGCCCATCTCCGCCCATCTCCGC 180
QY 181 CCCATCGCTGACTAAATTTTATTTATTCAGAGCGCGGCGCTCGGCTCTG 237
Db 181 CCCATCGCTGACTAAATTTTATTTATTCAGAGCGCGGCGCTCGGCTCTG 237
RESULT 2
AAZ11397 standard; DNA; 243 BP.
XX
XX AAZ11397;
XX
XX 26-OCT-1999 (first entry)
XX
XX Murine PGK fragment linked to SV40 promoter.
XX
XX Retroviral vector; functional splice donor site; hybrid viral vector;
KW functional splice acceptor site; in vivo gene delivery; therapeutic;
KW lentiviral vector; modified hematopoietic stem cell; MHSC; tumour;
KW ischaemia; hypoxia response element; HRE; hypoxia; ss.
XX
XX Synthetic.
OS Mus sp.
OS Rhesus macaque polyoma virus.
XX
XX WO9915684-A2.
XX
XX 01-APR-1999.
```

```
XX 23-SEP-1998; 98WO-GB02885.
PF
XX 25-SEP-1997; 97GB-0020465.
PR
XX 23-SEP-1997; 97GB-0020216.
XX
XX (OXFO-) OXFORD BIOMEDICA UK LTD.
XX
XX Bebbington C, Binley KM, Lewis C, Naylor S;
PI
XX WPI; 1999-263482/22.
DR
XX New retroviral vectors, for, e.g. delivering nucleotide sequences to
PT solid tumor sites
PT
XX Example 1; Page 69; 288pp; English.
XX
XX The invention relates to a retroviral vector (RVV) comprising a
CC functional splice donor site (FSDS) and a functional splice acceptor site
CC (FSAS) where: (i) the FSDS and the FSAS flank a first nucleotide sequence
CC of interest (NOI); (ii) the FSDS is upstream of the FSAS; (iii) the RVV
CC is derived from a retroviral pro-vector; (iv) the retroviral pro-vector
CC comprises a first nucleotide sequence (NS) capable of yielding the FSDS
CC and a second NS capable of yielding the FSAS; and (v) the first NS is
CC downstream of the second NS, such that the RVV is formed as a result of
CC reverse transcription of the retroviral pro-vector. A hybrid viral vector
CC (VV) system for in vivo gene delivery, which system comprises a primary
CC VV which encodes a secondary VV, the primary vector capable of infecting
CC a first target cell and of expressing the secondary VV, which secondary
CC vector is capable of transducing a secondary target cell, where the
CC primary vector is obtainable from or is based on a adenoviral vector and
CC the secondary VV is obtainable from or is based on a RVV preferably a
CC lentiviral vector (LVV) is also provided. The systems can be used for
CC delivering NOIs to one or more target sites. The NOIs may encode
CC therapeutic or diagnostic agents. The methods are used particularly for
CC producing modified hematopoietic stem cells (MHSCs) to deliver NOIs to
CC sites such as solid tumours which are characterised by ischemia, such as
CC hypoxia or low glucose concentration. The system permits the stable
CC expression of NOIs in targeted cells, e.g. rapidly dividing cells. The
CC present sequence represents a trimer encompassing -307/-290 sequence of
CC murine PGK, linked to SV40 promoter. The sequence comprises murine PGK
CC hypoxia response element (HRE) sequence.
XX
XX Sequence 243 BP; 45 A; 83 C; 58 G; 57 T; 0 other;
SQ
Query Match 93.2%; Score 221; DB 20; Length 243;
Best Local Similarity 97.5%; Pred. No. 5e-58;
Matches 237; Conservative 0; Mismatches 0; Indels 6; Gaps 1;
QY 1 GCTAGAGTCTGTCAGGACGTGACATCTAGTGTGTCGTCGAGG-----CATCTAGTGTCTG 54
Db 1 GCTAGAGTCTGTCAGGACGTGACATCTAGTGTGTCGTCGAGGACGTGACATCTAGTGTCTG 60
QY 55 CAGGACGTGACAGCTAGCCCGGCTCGAGATCTGCGATCTCATCTCAATTAGTCAGCAA 114
Db 61 CAGGACGTGACAGCTAGCCCGGCTCGAGATCTGCGATCTCATCTCAATTAGTCAGCAA 120
QY 115 CCATAGTCCGCGCCCTAACTCCGCCCATCCGCCCTCACTCCGCCAGTTCGCCCCATT 174
Db 121 CCATAGTCCGCGCCCTAACTCCGCCCATCCGCCCTCACTCCGCCAGTTCGCCCCATT 180
QY 175 CTCGCCCCCATCGCTGACTAAATTTTATTTATTCAGAGCGCGGCGCTCGGCT 234
Db 181 CTCGCCCCCATCGCTGACTAAATTTTATTTATTCAGAGCGCGGCGCTCGGCT 240
QY 235 CTG 237
Db 241 CTC 243
RESULT 3
AAZ11995
ID AAZ11995 standard; DNA; 243 BP.
```


agents. LD optimize activity of PAD, e.g. by delivering it to selected locations or by delivering it to neighboring cells (bystander effect), and allow a reduction in dose of prodrug, and thus of systemic side-effects. Nucleic acids encoding the agent may be expressed selectively in hypoxic cells. The present sequence represents a promoter OBHrel comprising a murine PGK fragment linked to SV40 promoter fragment. This is used in the EIAV vectors expressing P450.

XX Sequence 242 BP; 45 A; 83 C; 58 G; 56 T; 0 other;

Query Match 88.28; Score 209; DB 20; Length 242;
Best Local Similarity 97.18; Pred. No. 2.3e-54;
Matches 236; Conservative 0; Mismatches 0; Indels 7; Gaps 2;

QY 1 GCTAGAGTCGTGCAGGAGTGCATCTAGTCTGTCGTCAGG-----CATCTAGTGTGCTG 54
DB 1 GCTAGAGTCGTGCAGGAGTGCATCTAGTCTGTCGTCAGGAGTGCATCTAGTGTGCTG 60

QY 55 CAGGAGTCGACAGTCAGCCGGGCTCGAGATCTGGATCTGCATCTCAATTAGTCAGCAA 114
DB 61 CAGGAGTCGACAGTCAGCCGGGCTCGAGATCTGGATCTGCATCTCAATTAGTCAGCAA 120

QY 115 CCATAGTCCCGCCCTAACTCCGCCCATCCGCCCTAACTCCGCCCATTCGCCCCATT 174
DB 121 CCATAGTCCCGCCCTAACTCCGCCCATCCGCCCTAACTCCGCCCATTCGCCCCATT 179

QY 175 CTCGCCCCCATCGCTGACTAATTTTTTTTATTTATGTCAGAGGCGGCGGCTCGGCGCT 234
DB 180 CTCGCCCCCATCGCTGACTAATTTTTTTTATTTATGTCAGAGGCGGCGGCTCGGCGCT 239

QY 235 CTG 237
DB 240 CTG 242

RESULT 5
AA12016
ID AA12016 standard; DNA; 242 BP.
XX AC AA12016;
XX 14-AUG-2000 (first entry)
XX Murine PGK HRE derived promoter element DNA.
XX HRE; hypoxia response element; hypoxia-inducible factor; HIF; vasotropic;
KW cardiant; cytostatic; antiarthritic; gene therapy; ischaemia; arthritis;
KW cardiovascular disease; peripheral arterial disease; cancer;
KW phosphoglycerate kinase; PGK; murine; promoter; ds.
XX Mus sp.
XX WO200017371-A1.
XX 30-MAR-2000.
XX 22-SEP-1999; 99WO-GB03181.
XX 23-SEP-1998; 98WO-GB02885.
XX 28-JAN-1999; 99GB-0001906.
XX 16-FEB-1999; 99GB-0003538.
XX (OXFO-) OXFORD BIOMEDICA UK LTD.
XX Binley KM, Naylor S;
XX WPI; 2000-283595/24.
XX Novel polynucleotide constructs comprising at least two repeats of a
PT hypoxia response element useful for driving expression of nucleic acids
PT of interest in a cell under hypoxic conditions -
XX Example 1; Page 67; 155pp; English.

XX This invention describes novel polynucleotide comprising at least 2 repeats of a hypoxia response element (HRE), where the hypoxia-inducible factor (HIF) consensus binding sites within each of the 2 repeats are separated by a spacer of at least 20 contiguous nucleotides. The products of the invention have vasotropic, cardiant, cytostatic and antiarthritic activity and can be used for gene therapy. The polynucleotides are useful for delivering nucleic acids of interest to mammalian cells. Lentiviral vectors are responsive to hypoxic agents and to agents that mimic hypoxia. This regulation can be harnessed in vitro to enhance the production of the vector and can be used in vivo to regulate gene expression in response to a physiological signal. The vectors have utility in disease, where ischaemia, including hypoxia, is a feature, e.g. cardiovascular disease, peripheral arterial disease, cancer and arthritis. The novel regulatory construct is capable of driving very high levels of transcription under conditions of hypoxia whilst providing only low-basal levels of transcription under normal oxygen conditions. The polynucleotide construct targets cells within a tumor mass that are under conditions of hypoxia without affecting normal surrounding tissue. This sequence represents a murine phosphoglycerate kinase (PGK) HRE derived promoter element which is described in the method of the invention.

XX Sequence 242 BP; 45 A; 93 C; 47 G; 57 T; 0 other;

Query Match 73.8%; Score 175; DB 21; Length 242;
Best Local Similarity 84.8%; Pred. No. 5.9e-44;
Matches 196; Conservative 0; Mismatches 35; Indels 0; Gaps 0;

QY 7 GTCGTGCAGGAGTGCATCTAGTCTGTCGAGGCATCTAGTCTGTCGAGGAGTGCACA 66
DB 12 GTCCTGCAGCAGACTAGATGTCAGTCTCTGCACGACACTAGATGTCAGTCTGTCAGCAG 71

QY 67 GCTAGCCCGGCTCGAGATCTCGCATCTGCAATCTCAATTAGTCAGCAACCATGTCGCCG 126
DB 72 TCTAGCCCGGCTCGAGATCTCGCATCTGCAATCTCAATTAGTCAGCAACCATGTCGCCG 131

QY 127 CCTAACTCCGCCCATCCGCCCTAACTCCGCCCATCTCCGCCCATCTCCGCCCATC 186
DB 132 CCTAACTCCGCCCATCCGCCCTAACTCCGCCCATCTCCGCCCATCTCCGCCCATC 191

QY 187 GCTGACTAATTTTTTTTATTTATGTCAGAGGCGGCGGCTCGGCCCTCTG 237
DB 192 GCTGACTAATTTTTTTTATTTATGTCAGAGGCGGCGGCTCGGCCCTCTG 242

RESULT 6
AAD27539
ID AAD27539 standard; DNA; 5010 BP.
XX AC AAD27539;
XX 18-APR-2002 (first entry)
XX pGL3 promoter vector DNA.
XX p53 protein; pGL3 luciferase reporter vector; luc+; transcription factor;
KW cell cycle control; DNA damage repair; pGL3 promoter vector; apoptosis;
KW firefly; ds.
XX Chimeric - Photinus pyralis.
OS Chimeric - Unidentified.
XX Key Location/Qualifiers
FT misc_feature 1..41
FT /tag= a
FT /note= "Multiple cloning site"
FT promoter 49..244
FT /tag= b
FT misc_feature 280..1929
FT /tag= c
FT /note= "Luciferase gene (luc+)"
FT primer_bind complement (281..303)
FT /tag= d


```

FT polyA_signal /bound_moiety= "GL primer2"
FT 1964..2185 /tag= e
FT primer_bind /note= "SV40 late poly(A) signal"
FT Complement (2253..2272)
FT /tag= f
FT /bound_moiety= "RV primer4"
FT 2510
FT misc_signal /tag= g
FT /note= "ColE1-derived plasmid replication origin"
FT Complement (3275..4132)
FT /tag= h
FT misc_feature /note= "Beta-lactamase gene"
FT 4265..4719
FT /tag= i
FT /note= "F1 origin"
FT 4850..5003
FT polyA_signal /tag= j
FT /note= "Upstream poly(A) signal"
FT 4952..4971
FT primer_bind /tag= k
FT /bound_moiety= "RV primer3"
FT
FT WO200196602-A2.
FT
FT XX
FT XX
FT XX
FT XX
FT 18-JUN-2001; 2001WO-GB02718.
FT XX
FT 16-JUN-2000; 2000GB-0014820.
FT PR
FT (MEDI-) MEDICAL RES COUNCIL.
FT PA
FT Yang AL, Festing M;
FT PI
FT WPI; 2002-130743/17.
FT DR
FT
FT Determining the p53 status of a sample, useful for assaying for
FT mimetics or antagonists of p53, or for the presence of DNA damage,
FT comprises determining whether p53 binds to the pGL3 vector in a sample
FT containing a pGL3 vector -
FT
FT Claim 20; Page 42-44; 53pp; English.
FT
FT The patent discloses methods for determining the p53 status of a sample
FT which comprise providing a sample containing a pGL3 luciferase reporter
FT vector and determining whether p53 binds to the pGL3 vector. p53 is a
FT transcription factor that regulates many genes including those associated
FT with cell cycle control, apoptosis and DNA damage repair. pGL3 is a
FT vectors contain a modified firefly luciferase cDNA designated luc+. p53
FT protein binds to pGL3-basic vector and causes luciferase expression. The
FT method is useful for determining the p53 status of a sample. It is also
FT useful for assaying for mimetics or antagonists of p53 and for assaying
FT for presence of activated p53 and/or DNA damage. The invention also
FT relates to a method of modifying pGL3 vector which involves deletion
FT or alteration of a CCGGG motif of the pGL3 vector and/or deleting or
FT altering a sequence within 20 bp sequence 5' or 3' of CCGGG motif. The
FT nucleic acid having a sequence incorporating the CCGGG motif is useful
FT for conferring promoter activity or p53-responsiveness on a nucleic acid
FT encoding a polypeptide of interest or in assays for p53 transcriptional
FT activity. The present DNA sequence is pGL3 promoter vector sequence.
FT
FT SQ Sequence 5010 BP; 1287 A; 1213 C; 1198 G; 1312 T; 0 other;
FT
FT Query Match 72.2%; Score 171; DB 24; Length 5010;
FT Best Local Similarity 100.0%; Pred. No. 2.3e-42;
FT Matches 171; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
FT
FT QY 67 GCTAGCCGGGCTCGAGATCTGCGATCTCAATTAGTCAGCAACCATAGTCCCGC 126
FT |
FT Db 21 GCTAGCCGGGCTCGAGATCTGCGATCTCAATTAGTCAGCAACCATAGTCCCGC 80
FT |
FT QY 127 CCCTAACTCCGCCCATCCGCCCTTAAGTCGCGCCAGTTCGGCCCAATTCCTCCGCCCATC 186
FT |

```

```

Db 81 CCTAACTCCGCCCATCCGCCCTTAAGTCGCGCCAGTTCGGCCCAATTCCTCCGCCCATC 140
|
QY 187 GCTGACTAATTTTTTTTATTTATGTCAGAGCGGCGGCTCGGCGCTCTG 237
|
Db 141 GCTGACTAATTTTTTTTATTTATGTCAGAGCGGCGGCTCGGCGCTCTG 191
|
RESULT 7
AAA07776
ID AAA07776 standard; DNA; 5256 BP.
XX
AC AAA07776;
XX
DT 03-JUL-2000 (first entry)
XX
DE DNA sequence of plasmid pGL2.
XX
KW ced-6; hlced-6; h2ced-6; signal transduction pathway; phagocytosis;
KW cancer; autoimmune disease; neurodegenerative disease; stroke; AIDS;
KW Huntington's disease; myocardial infarction; cytostatic; neuroprotective;
KW cardiant; immunosuppressive; apoptosis modulator; luciferase; ss.
XX
OS Synthetic.
XX
PN WO9964586-A2.
XX
PD 16-DEC-1999.
XX
PF 10-JUN-1999; 99WO-EP04043.
XX
PR 11-JUN-1998; 98GB-0012660.
XX
PR 24-SEP-1998; 98GB-0020816.
XX
PA (DEVG-) DEVGEN NV.
XX
PI Smits E, Van Crielinge WMR, Bogaert TAOE;
XX
DR WPI; 2000-246285/21.
XX
PT Assays for determining the phagocytosis of apoptotic cells useful for
PT identifying a compound which influences the phagocytic uptake of
PT apoptotic cells and treats cancers and neurodegenerative diseases -
XX
PS Examples; Fig 19; 122pp; English.
XX
CC The invention relates to assays involving two human homologues of
CC Caenorhabditis elegans ced-6 (hlced-6 and h2ced-6) for identifying
CC compounds which function as an inhibitor or an enhancer of a signal
CC transduction pathway. The assays are carried out by measuring
CC phagocytosis of apoptotic cells. The methods are useful for identifying
CC compounds which can act as apoptotic modulators which are useful for
CC treating diseases such as cancer, autoimmune diseases, neurodegenerative
CC diseases such as Huntington's disease, stroke, myocardial infarction and
CC AIDS. The assays are well adapted for medium and high throughput
CC screening using a multi-well plate format. The present sequence
CC represents the DNA sequence of plasmid pGL2 which is suitable for
CC introduction of reporter gene luciferase into Ba/K3 cells.
XX
SQ Sequence 5256 BP; 1336 A; 1268 C; 1281 G; 1371 T; 0 other;
FT
FT Query Match 72.2%; Score 171; DB 21; Length 5256;
FT Best Local Similarity 100.0%; Pred. No. 2.3e-42;
FT Matches 171; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
FT
FT QY 67 GCTAGCCGGGCTCGAGATCTGCGATCTGCGATCTCAATTAGTCAGCAACCATAGTCCCGC 136
FT |
FT Db 21 GCTAGCCGGGCTCGAGATCTGCGATCTGCGATCTCAATTAGTCAGCAACCATAGTCCCGC 80
FT |
FT QY 127 CCCTAACTCCGCCCATCCGCCCTTAAGTCGCGCCAGTTCGGCCCAATTCCTCCGCCCATC 186
FT |
FT Db 81 CCCTAACTCCGCCCATCCGCCCTTAAGTCGCGCCAGTTCGGCCCAATTCCTCCGCCCATC 140
|

```

QY 187 GCTGACTAAATTTTATTATGACAGAGCCGAGCGCGCTCGGCTCTG 237
 DB 141 GCTGACTAAATTTTATTATGACAGAGCCGAGCGCGCTCGGCTCTG 191

RESULT 8

AD227537
 ID AD227537 standard; DNA; 5256 BP.
 XX
 AC AD227537;
 DT 18-APR-2002 (first entry)
 XX
 DE pGL3 control vector DNA.
 XX
 KW p53 protein; pGL3 luciferase reporter vector; luc+; transcription factor;
 KW cell cycle control; DNA damage repair; pGL3 control vector; apoptosis;
 KW firefly; ds.
 XX
 OS Chimeric - Photinus pyralis.
 OS Chimeric - Unidentified.
 XX
 FH Key Location/Qualifiers
 FT misc_feature 1..41
 FT /tag= a
 FT /note= "Multiple cloning site"
 FT promoter 49..244
 FT /tag= b
 FT misc_feature 280..1929
 FT /tag= c
 FT /note= "Luciferase gene (luc+)"
 FT primer_bind complement (281..303)
 FT /tag= d
 FT polyA_signal 1964..2185
 FT /bound_moiety= "GL primer2"
 FT /tag= e
 FT /note= "SV40 late poly(A) signal"
 FT enhancer 2197..2441
 FT /tag= f
 FT primer_bind complement (2499..2518)
 FT /tag= g
 FT /bound_moiety= "RV primer4"
 FT misc_signal 2756
 FT /tag= h
 FT /note= "ColE1-derived plasmid replication origin"
 FT misc_feature complement (3521..4378)
 FT /tag= i
 FT /note= "Beta-lactamase gene"
 FT misc_signal 4511..4965
 FT /tag= j
 FT /note= "F1 origin"
 FT polyA_signal 5096..5249
 FT /tag= k
 FT /note= "Upstream poly(A) signal"
 FT primer_bind 5198..5217
 FT /tag= l
 FT /bound_moiety= "RV primer3"
 FT
 FT
 XX WO200196602-A2.
 XX
 XX 20-DEC-2001.
 XX
 PF 18-JUN-2001; 2001WO-GB02718.
 XX
 XX 16-JUN-2000; 2000GB-0014820.
 XX
 PA (MEDI-) MEDICAL RES COUNCIL.
 XX
 XX Yang AL, Festing M;
 PI
 XX WPI; 2002-130743/17.
 DR
 XX Determining the p53 status of a sample, useful for assaying for

PT mimetics or antagonists of p53, or for the presence of DNA damage,
 PT comprises determining whether p53 binds to the pGL3 vector in a sample
 XX containing a pGL3 vector
 PS Claim 20; Page 36-39; 53pp; English.
 XX
 CC The patent discloses methods for determining the p53 status of a sample
 CC which comprise providing a sample containing a pGL3 luciferase reporter
 CC vector and determining whether p53 binds to the pGL3 vector. p53 is a
 CC transcription factor that regulates many genes including those associated
 CC with cell cycle control, apoptosis and DNA damage repair. pGL3 reporter
 CC vectors contain a modified firefly luciferase cDNA designated luc+. p53
 CC protein binds to pGL3-basic vector and causes luciferase expression. The
 CC method is useful for determining the p53 status of a sample. It is also
 CC useful for assaying for mimetics or antagonists of p53 and for assaying
 CC for presence of activated p53 and/or DNA damage. The invention also
 CC relates to a method of modifying pGL3 vector which involves deletion or
 CC alteration of a CCGGG motif of the pGL3 vector and/or deleting or
 CC altering a sequence within 20 bp sequence 5' or 3' of CCGGG motif. The
 CC nucleic acid having a sequence incorporating the CCGGG motif is useful
 CC for conferring promoter activity or p53-responsiveness on a nucleic acid
 CC encoding a polypeptide of interest or in assays for p53 transcriptional
 CC activity. The present DNA sequence is pGL3 control vector sequence.
 XX
 XX Sequence 5256 BP; 1336 A; 1268 C; 1281 G; 1371 T; 0 other;
 SQ
 Query Match 72.2%; Score 171; DB 24; Length 5256;
 Best Local Similarity 100.0%; Pred. No. 2.3e-42;
 Matches 171; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 67 GCTAGCCCGGGCTCGAGATCTGCGATCTGCAATTCATTTAGTCAGCAACCATAGTCCCGC 126
 DB 21 GCTAGCCCGGGCTCGAGATCTGCGATCTGCAATTCATTTAGTCAGCAACCATAGTCCCGC 80
 QY 127 CCTAACTCCGCCATCCGCCCTTAAGTCCGCCAGTTCGCCCATCTCTCCGCCCATC 186
 DB 81 CCTAACTCCGCCATCCGCCCTTAAGTCCGCCAGTTCGCCCATCTCTCCGCCCATC 140
 QY 187 GCTGACTAAATTTTATTATGACAGAGCCGAGCGCGCTCGGCTCTG 237
 DB 141 GCTGACTAAATTTTATTATGACAGAGCGGCGGCGCTCGGCTCTG 191
 RESULT 9
 AAAL2019
 ID AAAL2019 standard; DNA; 267 BP.
 XX
 AC AAAL2019;
 XX
 XX 14-AUG-2000 (first entry)
 XX
 XX Human erythropoietin enhancer Epo X4 DNA.
 DE
 DE HPE; hypoxia response element; hypoxia-inducible factor; HIF; vasotropic;
 KW cardiac; cytosolic; antiarthritic; gene therapy; ischaemia; arthritis;
 KW cardiovascular disease; peripheral arterial disease; cancer; human;
 KW enhancer; ds.
 XX
 XX Homo sapiens.
 XX
 XX WO200017371-A1.
 XX
 XX 30-MAR-2000.
 XX
 XX 22-SEP-1999; 99WO-GB03181.
 XX
 XX 23-SEP-1998; 98WO-GB02885.
 XX 28-JAN-1999; 99GB-0001906.
 XX 16-FEB-1999; 99GB-0003538.
 XX
 XX (OXFO-) OXFORD BIOMEDICA UK LTD.
 XX
 XX Binley KM, Naylor S;
 PI

XX Homo sapiens.
 OS WO200017371-A1.
 PN 30-MAR-2000.
 XX 22-SEP-1999; 99WO-GB03181.
 XX 23-SEP-1998; 98WO-GB02885.
 PR 28-JAN-1999; 99GB-0001906.
 PR 16-FEB-1999; 99GB-0003538.
 XX (OXFO-) OXFORD BIOMEDICA UK LTD.
 XX Binley KM, Naylor S;
 PI WPI; 2000-283595/24.
 DR Novel polynucleotide constructs comprising at least two repeats of a
 PT hypoxia response element useful for driving expression of nucleic acids
 PT of interest in a cell under hypoxic conditions -
 XX Example 1; Page 69; 155pp; English.
 XX This invention describes novel polynucleotide comprising at least 2
 CC repeats of a hypoxia response element (HRE), where the hypoxia-inducible
 CC factor (HIF) consensus binding sites within each of the 2 repeats are
 CC separated by a spacer of at least 20 contiguous nucleotides. The products
 CC of the invention have vasotropic, cardiant, cytostatic and antiarthritic
 CC activity and can be used for gene therapy. The polynucleotides are useful
 CC for delivering nucleic acids of interest to mammalian cells. Lentiviral
 CC vectors are responsive to hypoxic agents and to agents that mimic
 CC hypoxia. This regulation can be harnessed in vitro to enhance the
 CC production of the vector and can be used in vivo to regulate gene
 CC expression in response to a physiological signal. The vectors have
 CC utility in disease, where ischaemia, including hypoxia, is a feature,
 CC e.g. cardiovascular disease, peripheral arterial disease, cancer and
 CC arthritis. The novel regulatory construct is capable of driving very high
 CC levels of transcription under conditions of hypoxia whilst providing only
 CC low basal levels of transcription under normal oxygen conditions. The
 CC polynucleotide construct targets cells within a tumor mass that are under
 CC conditions of hypoxia without affecting normal surrounding tissue. This
 CC sequence represents a murine phosphoglycerate kinase (PGK) HRE derived
 CC promoter element which is described in the method of the invention.
 XX Sequence 223 BP; 38 A; 81 C; 58 G; 46 T; 0 other;
 SQ
 Query Match 62.4%; Score 148; DB 21; Length 223;
 Best Local Similarity 100.0%; Pred. No. 1e-35;
 Matches 148; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 90 GATCTGATCTCAATTAGTCAGCAACCATAGTCCGCCCTTACTCCGCCCATCCGCC 149
 DB 76 GATCTGATCTCAATTAGTCAGCAACCATAGTCCGCCCTTACTCCGCCCATCCGCC 135
 QY 150 CTACTCGCCCGAGTTCGCCCATCTTCCGCCCATCGCTGACTAATTTTTTTTAT 209
 DB 136 CTACTCGCCCGAGTTCGCCCATCTTCCGCCCATCGCTGACTAATTTTTTTTAT 195
 QY 210 GCAGAGGCGGAGCGCGCTCGGCTCTG 237
 DB 196 GCAGAGGCGGAGCGCGCTCGGCTCTG 223
 RESULT 12
 ID AAT42469
 XX AAT42469 standard; cDNA; 356 BP.
 AC AAT42469;
 XX 05-AUG-1997 (first entry)
 DT
 XX

DE Interleukin-6 inhibitor BamHI-HindIII fragment of pm8SV.
 XX inhibitor; interleukin-6; IL-6; acute phase response element; APRE;
 KW transcription factor binding site; therapy; autoimmune; Castleman's;
 KW inflammatory disease; rheumatoid arthritis; glomerulonephritis;
 KW psoriasis; ss.
 XX Synthetic.
 OS
 XX Key Location/Qualifiers
 FH 11..155
 FT misc_feature /*tag= a
 FT /*note= "M8 sequence"
 FT 11..19
 FT misc_signal /*tag= b
 FT /*note= "acute phase response element; claim 2"
 FT 160..356
 FT misc_feature /*tag= c
 FT /*note= "SV40 promoter sequence"
 XX WO9635782-A1.
 PN 14-NOV-1996.
 XX 11-MAY-1995; 95WO-EP01778.
 XX 11-MAY-1995; 95WO-EP01778.
 XX (ISTF) ARS APPLIED RES SYST HOLDING NV.
 PA Pezzotti A, Serlupi-Crescenzi O;
 PI WPI; 1996-518670/51.
 DR Nucleotide sequence which inhibits interleukin-6 activity - used in
 XX pharmaceutical compositions for treatment of auto-immune and
 PT inflammatory diseases
 PT
 XX Claim 10; Fig 2; 35pp; English.
 XX The present sequence comprises the BamHI-HindIII fragment of pm8SVL
 CC which contains a nucleotide sequence which is able to inhibit
 CC interleukin-6 (IL-6) activity. The nucleotide sequence contains an acute
 CC phase response element (APRE) and at least one nucleotide sequence
 CC constituting a transcription factor binding site other than the APRE
 CC element. The nucleotide sequences can be used in therapy to inhibit the
 CC action of IL-6 in conditions where IL-6 plays a pathological role, e.g.
 CC autoimmune and inflammatory diseases such as rheumatoid arthritis,
 CC psoriasis, Castleman's disease and glomerulonephritis. Other methods of
 CC inhibiting IL-6 activity, such as use of antibodies against IL-6, gp130
 CC or gp80, or the use of mutants for IL-6 or the IL-6 receptor, may be
 CC associated with undesired clinical effects. The new approach involves
 CC blocking the intracellular proteins mediating the IL-6 signal.
 XX Sequence 356 BP; 66 A; 103 C; 76 G; 111 T; 0 other;
 SQ
 Query Match 62.2%; Score 147.4; DB 17; Length 356;
 Best Local Similarity 96.2%; Pred. No. 1.8e-35;
 Matches 151; Conservative 0; Mismatches 6; Indels 0; Gaps 0;
 QY 81 GAGATCTGCGATCTGATCTCAATTAGTCAGCAACCATAGTCCGCCCTTACTCCGCC 140
 DB 141 GGGAAATTCATGATCTGATCTCAATTAGTCAGCAACCATAGTCCGCCCTTACTCCGCC 200
 QY 141 ATCCCGCCCTTACTCGGCCAGTTCGCCCATCTTCCGCCCATCGCTGACTAATTTTT 200
 DB 201 ATCCCGCCCTTACTCGGCCAGTTCGCCCATCTTCCGCCCATCGCTGACTAATTTTT 260
 QY 201 TTTATTTATGAGAGGCGGAGCGCGCTCGGCTCTG 237
 DB 261 TTTATTTATGAGAGGCGGAGCGCGCTCGGCTCTG 297

RESULT 13

AXX08776
ID AAX08776 standard; DNA; 5793 BP.

AA
AC

XX
DT 27-SEP-1999 (first entry)

pGL2(apo AI-ARE)SV40/luc plasmid construct comprising ARE/luciferase.

Antioxidant responsive element; ARE; low density lipoprotein; LDL; high density lipoprotein; HDL; apolipoprotein; apo AI; atherosclerosis; heart disease; transcription; ss.

Synthetic.

XX
PN CA2238662-A.

23-NOV-1998.

XX
PF
22-MAY-1998; 98CA-2238662.

23-MAY-1997; 97US-0862431

(TOOH) UNIV QUEENS KINGSTON.

Tam S:

WPI; 1999-229918/20.

New Antioxidant Response Element (ARE), useful for identifying drugs and transcription factors for increasing transcription of mRNA, useful for treatment of atherosclerosis

S Examples; Page 63-66; 115pp; English.

DNA constructs comprising antioxidant responsive elements (AREs) are useful for screening for compounds and transcription factors that bind to the ARE and increase transcription levels of a mRNA regulated by an ARE. AREs may also be useful as a reagent for purification of a compound (preferably a transcription factor) with which it interacts. High Density Lipoprotein (HDL) has antioxidative activity and protects against oxidized low-density lipoprotein (LDL) which has a role in the etiology of atherosclerosis. Apolipoprotein (apo) AI is a major component of HDL, and is believed to promote the process of reverse cholesterol transport. The transcription of apo AI is effected by cis- and trans-acting factors (i.e. an ARE) UV cross-linking studies using an apoAI-ARE probe isolated two polypeptides of 100 and 115 kDa. These compounds are useful for treatment of a human or animal with atherosclerosis. ARE's can also be used in DNA constructs when operably linked to heterologous protein coding sequences to effect the transcription of those heterologous sequences. A Genelight vector (pGL2-B) comprising a 491 bp fragment of the apo AI from -491 to +1 inserted into the XhoI site of the vector and upstream of the luciferase gene was designated pGL2(apoAI-491)luc. The plasmid pGL2(apoAI-250)luc was constructed by replacing a DNA fragment (-491 to -251 of the apo AI promoter) from pGL2(apoAI-491)luc. A new plasmid pGL(apoAI-250 mutant ARE) was prepared by PCR using four primers (AAX08785-88). Two primers (AAX08785, AAX08786) were complementary to the non-coding strand of the remaining two primers (AAX08787, AAX08788) were complementary to the coding strand of DNA. The plasmid pGL2(apo-AI250)luc was used as a template and the DNA fragment was purified and then cloned into the SmaI and HindIII sites of pGL2-B to generate pGL2(apoAI-250 mutant ARE). A series of pGL2-Promoter vectors comprising apoAI-ARE, GST-ARE and mutated ARE were also constructed. See AAX08773-X08779.

Sequence 5793 BP; 1562 A; 1330 C; 1310 G; 1591 T; 0 other;

Query Match 62.0%; Score 147; DB 20; Length 5793;
Best Local Similarity 86.6%; Pred. No. 5.3e-35;

[illegible]

RESULT 14

MAX

ID AAX08779 standard; DNA; 5789 BP.

XX AAX08779;

DT 27-SEP-1999 (first entry)

xx
DE
yv
pGL2-Promoter genelight vector comprising ARE and apo-AI promoter.

Antioxidant responsive element; ARE; low density lipoprotein; LDL;
high density lipoprotein; HDL; apolipoprotein; apo AI;
atherosclerosis; heart disease; transcription; ss.

OS Synthetic.

XX
PN CA2238662XX
PD 23-NOV-1998.XX
PF
22-MAY-1998; 98CA-2238662.

23-MAY-1997; 97US-0862431.

(TOOH) UNIV QUEENS KINGSTON.

Tam S:

WPI; 1999-229918/20.

New Antioxidant Response Element (ARE), useful for identifying drugs and transcription factors for increasing transcription of mRNA, useful for treatment of atherosclerosis

Examples; Page 74-77; 115pp; English.

DNA constructs comprising antioxidant responsive elements (AREs) are useful for screening for compounds and transcription factors that bind to the ARE and increase transcription levels of a mRNA regulated by an ARE. AREs may also be useful as a reagent for purification of a compound (preferably a transcription factor) with which it interacts. High Density Lipoprotein (HDL) has antioxidant activity and protects against oxidized low-density lipoprotein (LDL) which has a role in the etiology of atherosclerosis. Apolipoprotein (apo) AI is a major component of HDL, and is believed to promote the process of reverse cholesterol transport. The transcription of apo AI is effected by cis- and trans-acting factors (i.e. an ARE) UV cross-linking studies using an apoAI-ARE probe isolated two polypeptides of 100 and 115 kDa. These compounds are useful for treatment of a human or animal with atherosclerosis. ARE's can also be used in DNA constructs when operably linked to heterologous protein coding sequences to effect the transcription of those heterologous sequences. A Genesight vector (pGL2-B) comprising a 491 bp fragment of the apo

regulated by an ARE. AREs may also be useful as a reagent for purification of a compound (preferably a transcription factor) with which it interacts. High Density Lipoprotein (HDL) has antioxidant activity and protects against oxidized low-density lipoprotein (LDL) which has a role in the etiology of atherosclerosis. Apolipoprotein (apo) AI is a major component of HDL, and is believed to promote the process of reverse cholesterol transport. The transcription of apo AI is effected by cis- and trans-acting factors (i.e. an ARE) UV cross-linking studies using an apoAI-ARE probe isolated two polypeptides of 100 and 115 kDa. These compounds are useful for treatment of a human or animal with atherosclerosis. ARE's can also be used in DNA constructs when operably linked to heterologous protein coding sequences. A effect the transcription of those heterologous sequences. A gene/light vector (pGL2-B) comprising a 491 bp fragment of the apo AI from -491 to +1 inserted into the XhoI site of the vector and upstream of the luciferase gene was designated pGL2(apoAI-491)luc. Plasmid pGL2(apoAI-250)luc was constructed by releasing a DNA fragment (-491 to -251 of the apo AI promoter) from pGL2(apoAI-491)luc. A new plasmid pGL2(apoAI-250 mutant ARE) was prepared by PCR using four primers (AAx08785-88). Two primers (AAx08785, AAx08786) were complementary to the non-coding strand of DNA, the remaining two primers (AAx08787, AAx08788) were complementary to the coding strand of DNA. The plasmid pGL2(apoAI-250)luc was used as a template and the DNA fragment was purified and then cloned into the SmaI and HindIII sites of pGL2-B to generate pGL2(apoAI-250 mutant ARE). A series of pGL2-Promoter vectors comprising apoAI-ARE, GST-ARE and mutated ARE were also constructed. See AAx08773-X08779.

Sequence 5789 BP; 1559 A; 1328 C; 1308 G; 1594 T; 0 other;

Query Match 61.9%; Score 146.8; DB 20; Length 5789;
Best Local Similarity 98.7%; Pred. No. 6.1e-35;
Matches 148; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 88 GCGATCTGCATCTCAATTAGTCAGCAACCATAGTCCCGCCCTAACTCCGCCCATCCCGC 147
DB 36 GAGATCTGCATCTCAATTAGTCAGCAACCATAGTCCCGCCCTAACTCCGCCCATCCCGC 95
QY 148 CCCTAACTCCGCCCATCTCCGCCCATCTCCGCCCATCTCCGCCCATCTCCGCCCATCT 207
DB 96 CCCTAACTCCGCCCATCTCCGCCCATCTCCGCCCATCTCCGCCCATCTCCGCCCATCT 155
QY 208 ATGCAGAGCCGAGGCCGCTCGGCTCTG 237
DB 156 ATGCAGAGCCGAGGCCGCTCGGCTCTG 185

RESULT 15
AAx08777
ID AAx08777 standard; DNA; 5793 BP.
AC AAx08777;
XX
XX 27-SEP-1999 (first entry)
XX pGL2(apo AI-mARE)SV40/luc construct comprising ARE/luciferase.
DE Antioxidant responsive element; ARE; low density lipoprotein; LDL;
KW high density lipoprotein; HDL; apolipoprotein; apo AI;
KW atherosclerosis; heart disease; transcription; ss.
XX Synthetic.
XX
XX CA2238662-A.
XX
XX 23-NOV-1998.
XX
XX 22-MAY-1998; 98CA-2238662.
XX
XX 23-MAY-1997; 97US-0862431.
XX
XX (TOOH) UNIV QUEENS KINGSTON.
XX
XX Tam S;
XX
XX WPI; 1999-229918/20.
XX
XX New Antioxidant Response Element (ARE), useful for identifying drugs
PT and transcription factors for increasing transcription of mRNA,
PT useful for treatment of atherosclerosis
XX
XX Examples; Page 66-70; 115pp; English.
PS
XX DNA constructs comprising antioxidant responsive elements (AREs)
CC are useful for screening for compounds and transcription factors
CC that bind to the ARE and increase transcription levels of a mRNA

regulated by an ARE. AREs may also be useful as a reagent for purification of a compound (preferably a transcription factor) with which it interacts. High Density Lipoprotein (HDL) has antioxidant activity and protects against oxidized low-density lipoprotein (LDL) which has a role in the etiology of atherosclerosis. Apolipoprotein (apo) AI is a major component of HDL, and is believed to promote the process of reverse cholesterol transport. The transcription of apo AI is effected by cis- and trans-acting factors (i.e. an ARE) UV cross-linking studies using an apoAI-ARE probe isolated two polypeptides of 100 and 115 kDa. These compounds are useful for treatment of a human or animal with atherosclerosis. ARE's can also be used in DNA constructs when operably linked to heterologous protein coding sequences. A effect the transcription of those heterologous sequences. A gene/light vector (pGL2-B) comprising a 491 bp fragment of the apo AI from -491 to +1 inserted into the XhoI site of the vector and upstream of the luciferase gene was designated pGL2(apoAI-491)luc. Plasmid pGL2(apoAI-250)luc was constructed by releasing a DNA fragment (-491 to -251 of the apo AI promoter) from pGL2(apoAI-491)luc. A new plasmid pGL2(apoAI-250 mutant ARE) was prepared by PCR using four primers (AAx08785-88). Two primers (AAx08785, AAx08786) were complementary to the non-coding strand of DNA, the remaining two primers (AAx08787, AAx08788) were complementary to the coding strand of DNA. The plasmid pGL2(apoAI-250)luc was used as a template and the DNA fragment was purified and then cloned into the SmaI and HindIII sites of pGL2-B to generate pGL2(apoAI-250 mutant ARE). A series of pGL2-Promoter vectors comprising apoAI-ARE, GST-ARE and mutated ARE were also constructed. See AAx08773-X08779.

Sequence 5793 BP; 1561 A; 1328 C; 1308 G; 1596 T; 0 other;

Query Match 61.9%; Score 146.8; DB 20; Length 5793;
Best Local Similarity 98.7%; Pred. No. 6.1e-35;
Matches 148; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 88 GCGATCTGCATCTCAATTAGTCAGCAACCATAGTCCCGCCCTAACTCCGCCCATCCCGC 147
DB 40 GAGATCTGCATCTCAATTAGTCAGCAACCATAGTCCCGCCCTAACTCCGCCCATCCCGC 99
QY 148 CCCTAACTCCGCCCATCTCCGCCCATCTCCGCCCATCTCCGCCCATCTCCGCCCATCT 207
DB 100 CCCTAACTCCGCCCATCTCCGCCCATCTCCGCCCATCTCCGCCCATCTCCGCCCATCT 159
QY 208 ATGCAGAGCCGAGGCCGCTCGGCTCTG 237
DB 160 ATGCAGAGCCGAGGCCGCTCGGCTCTG 189

Search completed: January 3, 2003, 23:20:43
Job time : 398.246 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM.nucleic - nucleic search, using sw model

Run on: January 3, 2003, 22:53:16 ; Search time 79.9968 Seconds
(without alignments)
908.566 Million cell updates/sec

Title: US-09-787-562-9
Perfect score: 237
Sequence: 1 gctagatctgctgagacgt.....cgaggccgctcggtctg 237
Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0
Searched: 441362 seqs, 153338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NA: *
1: /cgn2_6/ptodata/2/ina/5A_COMB.seq: *
2: /cgn2_6/ptodata/2/ina/5B_COMB.seq: *
3: /cgn2_6/ptodata/2/ina/6A_COMB.seq: *
4: /cgn2_6/ptodata/2/ina/6B_COMB.seq: *
5: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq: *
6: /cgn2_6/ptodata/2/ina/backfiles1.seq: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	147.4	62.2	356	3	US-08-945-726-1
2	147	62.0	5793	3	US-08-862-431-29
3	146.8	61.9	5789	3	US-08-862-431-32
4	146.8	61.9	5791	3	US-08-862-431-31
5	146.8	61.9	5793	3	US-08-862-431-30
6	145.2	61.3	7731	4	US-09-301-593-29
7	145.2	61.3	7731	4	US-09-301-593-42
8	142.8	60.3	8068	4	US-09-301-593-27
9	142.8	60.3	8068	4	US-09-301-593-35
10	142.8	60.3	13254	1	US-08-276-852-156
11	142.8	60.3	13254	1	US-08-276-852-170
12	142.8	60.3	13254	1	US-08-899-575-156
13	142.8	60.3	13254	1	US-08-899-575-170
14	142.8	60.3	13254	1	US-08-899-575-156
15	142.8	60.3	13254	1	US-08-899-575-170
16	142.8	60.3	13254	5	PCT-US95-08743-156
17	142.8	60.3	13254	5	PCT-US95-08743-170
18	142.4	60.1	372	1	US-08-822-982-1
19	142.4	60.1	600	5	PCT-US91-02766-23
20	142.4	60.1	600	6	5219759-5
21	142.4	60.1	633	1	US-08-644-664B-6
22	142.4	60.1	633	2	US-08-761-277A-6
23	142.4	60.1	664	1	US-08-455-755-1
24	142.4	60.1	664	2	US-07-910-260-1
25	142.4	60.1	664	5	PCT-US91-00878-1
26	142.4	60.1	2800	3	US-08-889-841B-42
27	142.4	60.1	3705	3	US-09-282-996-3

28	142.4	60.1	3740	4	US-08-955-565A-6	Sequence 6, Appli
29	142.4	60.1	4279	4	US-09-446-402A-1	Sequence 1, Appli
30	142.4	60.1	4454	1	US-07-712-284-1	Sequence 1, Appli
31	142.4	60.1	4454	5	PCT-US92-04227-1	Sequence 1, Appli
32	142.4	60.1	4518	4	US-09-380-190A-26	Sequence 26, Appli
33	142.4	60.1	4627	4	US-09-054-281-19	Sequence 19, Appli
34	142.4	60.1	4644	4	US-09-479-122-19	Sequence 19, Appli
35	142.4	60.1	4665	3	US-08-948-378A-7	Sequence 7, Appli
36	142.4	60.1	4665	4	US-09-169-425C-7	Sequence 7, Appli
37	142.4	60.1	4696	2	US-08-929-967-15	Sequence 15, Appli
38	142.4	60.1	4732	1	US-07-884-811-1	Sequence 1, Appli
39	142.4	60.1	4732	1	US-07-885-971-1	Sequence 1, Appli
40	142.4	60.1	4732	1	US-08-087-783A-1	Sequence 1, Appli
41	142.4	60.1	4732	1	US-08-194-088B-1	Sequence 1, Appli
42	142.4	60.1	4732	2	US-08-194-087-1	Sequence 1, Appli
43	142.4	60.1	4732	5	PCT-US93-04648-1	Sequence 1, Appli
44	142.4	60.1	4886	4	US-09-533-220A-4	Sequence 4, Appli
45	142.4	60.1	4951	2	US-08-752-307B-1	Sequence 1, Appli

ALIGNMENTS

RESULT 1
US-08-945-726-1
; Sequence 1, Application US/08945726
; Patent No. 6004813
; GENERAL INFORMATION:
; APPLICANT: SERLUPI-CRESCENZI, Ottaviano
; APPLICANT: PEZZOTTI, Annarita
; TITLE OF INVENTION: IL-6 ACTIVITY INHIBITORS
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: BROWDY AND NEIMARK, P.L.L.C.
; STREET: 419 Seventh Street, N.W., Suite 300
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20004
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/945.726
; FILING DATE: 10-NOV-1997
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: BROWDY, Roger L.
; REGISTRATION NUMBER: 25,618
; REFERENCE/DOCKET NUMBER: SERLUPI-1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-628-5197
; TELEFAX: 202-737-3528
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 356 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: CDNA
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
US-08-945-726-1

Query Match 62.2%; Score 147.4; DB 3; Length 356;
Best Local Similarity 96.2%; Pred. No. 4.9e-36;
Matches 151; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 81 GAGATCTGCGATCTGCAATTCATAGTCAGCAACCATAGTCCGCCCTAACTCGGCC 140
DB 141 GGGATCTGATCTGCAATTCATAGTCAGCAACCATAGTCCGCCCTAACTCGGCC 200

Patent No. 6120994
GENERAL INFORMATION:
APPLICANT: TAM, SHUI-PANG
TITLE OF INVENTION: ANTIOXIDANT RESPONSIVE ELEMENT
NUMBER OF SEQUENCES: 51
CORRESPONDENCE ADDRESS:
ADDRESSEE: STERNE, KESSLER, GOLDSTEIN & FOX P.L.L.C.
STREET: 1100 NEW YORK AVENUE, SUITE 600
CITY: WASHINGTON
STATE: DC
COUNTRY: US
ZIP: 20005-3934
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/862,431
FILING DATE: 23-MAY-1997
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Kim, Judith U.
REGISTRATION NUMBER: 40,679
REFERENCE/DOCKET NUMBER: 1669.0020000
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 371-2600
TELEFAX: (202) 371-2540
INFORMATION FOR SEQ ID NO: 32:
SEQUENCE CHARACTERISTICS:
LENGTH: 5789 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: circular
US-08-862-431-32

Query Match 61.9%; Score 146.8; DB 3; Length 5789;
Best Local Similarity 98.7%; Pred. No. 1.8e-35;
Matches 148; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 88 GCGATCTGCATCTCAATTAGTCAGCAACCATAGTCCGCCCTAACTCCGCCCATCCGC 147
DB 36 GAGATCTGCATCTCAATTAGTCAGCAACCATAGTCCGCCCTAACTCCGCCCATCCGC 95
QY 148 CCCTAACTCCGCCCATCTCCGCCCATCTCCGCCCATCTCCGCCCATCTCCGCCCATCT 207
DB 96 CCCTAACTCCGCCCATCTCCGCCCATCTCCGCCCATCTCCGCCCATCTCCGCCCATCT 155

QY 208 ATGCAGAGCCGAGCGCGCTCGGCGCTCTG 237
DB 156 ATGCAGAGCCGAGCGCGCTCGGCGCTCTG 185

RESULT 4
US-08-862-431-31
Sequence 31, Application US/08862431
Patent No. 6120994
GENERAL INFORMATION:
APPLICANT: TAM, SHUI-PANG
TITLE OF INVENTION: ANTIOXIDANT RESPONSIVE ELEMENT
NUMBER OF SEQUENCES: 51
CORRESPONDENCE ADDRESS:
ADDRESSEE: STERNE, KESSLER, GOLDSTEIN & FOX P.L.L.C.
STREET: 1100 NEW YORK AVENUE, SUITE 600
CITY: WASHINGTON
STATE: DC
COUNTRY: US
ZIP: 20005-3934
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/862,431
FILING DATE: 23-MAY-1997
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Kim, Judith U.
REGISTRATION NUMBER: 40,679
REFERENCE/DOCKET NUMBER: 1669.0020000
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 371-2600
TELEFAX: (202) 371-2540
INFORMATION FOR SEQ ID NO: 29:
SEQUENCE CHARACTERISTICS:
LENGTH: 5793 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: circular
US-08-862-431-29

Query Match 62.0%; Score 147; DB 3; Length 5793;
Best Local Similarity 86.6%; Pred. No. 1.5e-35;
Matches 162; Conservative 0; Mismatches 25; Indels 0; Gaps 0;

QY 51 GGTGAGAGCTGACAGCTAGCGCGGCTCGAGATCTGCGATCTGCAATCTCAATTAGTCA 110
DB 3 CGGAGGTACAGCCCGGAGGACAGAGCTGCTAGTCTGAGATCTGCAATCTCAATTAGTCA 62
QY 111 GCAACCATAGTCCCGCCCTAACTCCGCCCATCTCCGCCCATCTCCGCCCATCTCCGCC 170
DB 63 GCAACCATAGTCCCGCCCTAACTCCGCCCATCTCCGCCCATCTCCGCCCATCTCCGCC 122
QY 171 CATCTCCGCCCATCTGCTGACTAATTTTTTTTATTTATTCAGAGGCGGAGCGGCTCG 230
DB 123 CATCTCCGCCCATCTGCTGACTAATTTTTTTTATTTATTCAGAGGCGGAGCGGCTCG 182

QY 231 GCCTCTG 237
DB 183 GCCTCTG 189

RESULT 3
US-08-862-431-32
Sequence 32, Application US/08862431

Query Match 86.6%; Score 147; DB 3; Length 5793;
Best Local Similarity 86.6%; Pred. No. 1.5e-35;
Matches 162; Conservative 0; Mismatches 25; Indels 0; Gaps 0;

QY 51 GGTGAGAGCTGACAGCTAGCGCGGCTCGAGATCTGCGATCTGCAATCTCAATTAGTCA 110
DB 3 CGGAGGTACAGCCCGGAGGACAGAGCTGCTAGTCTGAGATCTGCAATCTCAATTAGTCA 62
QY 111 GCAACCATAGTCCCGCCCTAACTCCGCCCATCTCCGCCCATCTCCGCCCATCTCCGCC 170
DB 63 GCAACCATAGTCCCGCCCTAACTCCGCCCATCTCCGCCCATCTCCGCCCATCTCCGCC 122
QY 171 CATCTCCGCCCATCTGCTGACTAATTTTTTTTATTTATTCAGAGGCGGAGCGGCTCG 230
DB 123 CATCTCCGCCCATCTGCTGACTAATTTTTTTTATTTATTCAGAGGCGGAGCGGCTCG 182

QY 231 GCCTCTG 237
DB 183 GCCTCTG 189

RESULT 5
 US-08-862-431-30
 ; Sequence 30, Application US/08862431
 ; Patent No. 6120994
 ; GENERAL INFORMATION:
 ; APPLICANT: TAM, SHUI-PANG
 ; TITLE OF INVENTION: ANTIOXIDANT RESPONSIVE ELEMENT
 ; NUMBER OF SEQUENCES: 51
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: STERNE, KESSLER, GOLDSTEIN & FOX P.L.L.C.
 ; STREET: 1100 NEW YORK AVENUE, SUITE 600
 ; CITY: WASHINGTON
 ; STATE: DC
 ; COUNTRY: US
 ; ZIP: 20005-3934
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: PatentIn Release #1.0, Version #1.30
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/862,431
 ; FILING DATE: 23-MAY-1997
 ; CLASSIFICATION: 435
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Kim, Judith U.
 ; REGISTRATION NUMBER: 40,679
 ; REFERENCE/DOCKET NUMBER: 1669.0020000
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (202) 371-2600
 ; TELEFAX: (202) 371-2540
 ; INFORMATION FOR SEQ ID NO: 30:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 5793 base pairs
 ; TYPE: nucleic acid
 ; STRANDEDNESS: double
 ; TOPOLOGY: circular

RESULT 7
US-09-301-593-42
; Sequence 42, Application US/09301593A
; Patent No. 6455677
; GENERAL INFORMATION:
; APPLICANT: Park, John E.

;; TITLE OF INVENTION: HUMAN NEUTRALIZING MONOCLONAL ANTIBODIES
;; TITLE OF INVENTION: TO HUMAN IMMUNODEFICIENCY VIRUS
;; NUMBER OF SEQUENCES: 170
;; CORRESPONDENCE ADDRESS:
;; ADDRESSEE: The Scripps Research Institute, Office of
;; ADDRESSEE: Patent Counsel
;; STREET: 10666 No. 5652138th Torrey Pines Road, Suite 220,
;; STREET: Mail Drop TPC8
;; CITY: La Jolla
;; STATE: CA
;; COUNTRY: USA
;; ZIP: 92037
;; COMPUTER READABLE FORM:
;; MEDIUM TYPE: Floppy disk
;; COMPUTER: IBM PC compatible
;; OPERATING SYSTEM: PC-DOS/MS-DOS
;; SOFTWARE: Patentin Release #1.0, Version #1.25
;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/08/276,852
;; FILING DATE: 18-JUL-1994
;; CLASSIFICATION: 514
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: US 08/178,302
;; FILING DATE: 30-SEP-1993
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: US 07/954,148
;; FILING DATE: 30-SEP-1992
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Fitting, Thomas
;; REGISTRATION NUMBER: 34,163
;; REFERENCE/DOCKET NUMBER: SCR1452P
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: 619-554-2937
;; TELEFAX: 619-554-6312
;; INFORMATION FOR SEQ ID NO: 156:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 13254 base pairs
;; TYPE: nucleic acid
;; STRANDEDNESS: double
;; TOPOLOGY: circular
;; MOLECULE TYPE: DNA (genomic)
;; US-08-276-852-156

Query Match 60.3%; Score 142.8; DB 1; Length 13254;
Best Local Similarity 81.7%; Pred. No. 3.7e-34;
Matches 165; Conservative 0; Mismatches 37; Indels 0; Gaps 0;
QY 36 GCAGGCATCTAGTGTGTCGAGGACGTGACAGCTAGCCCGGCTCGAGATCTGGGATCTG 95
Db 8076 GCATGATCTCAATTAGTCAGCAACAGGCTCCCGCAGCAGCAAGTATGCAAGCATG 8135
QY 96 CATCTCAATTAGTCAGCAACCATAGTCCCGCCCTAACTCGGCCCATCCGCCCTAACT 155
Db 8136 CATCTCAATTAGTCAGCAACCATAGTCCCGCCCTAACTCGGCCCATCCGCCCTAACT 8195
QY 156 CGGCCAGTTCGGCCATCTCCGCCCATCGCTGACTAATTTTTTTTATTTATGAGAG 215
Db 8196 CGGCCAGTTCGGCCATCTCCGCCCATCGCTGACTAATTTTTTTTATTTATGAGAG 8255
QY 216 GCCGAGGCGGCTCGGCTCTG 237
Db 8256 GCCGAGGCGGCTCGGCTCTG 8277

RESULT 11
US-08-276-852-170/c
; Sequence 170, Application US/08276852
; Patent No. 5652138
; GENERAL INFORMATION:
; APPLICANT: Burton, Dennis R
; APPLICANT: Barbas, Carlos F
; APPLICANT: Lerner, Richard A
; TITLE OF INVENTION: HUMAN NEUTRALIZING MONOCLONAL ANTIBODIES

;; TITLE OF INVENTION: TO HUMAN IMMUNODEFICIENCY VIRUS
;; NUMBER OF SEQUENCES: 170
;; CORRESPONDENCE ADDRESS:
;; ADDRESSEE: The Scripps Research Institute, Office of
;; ADDRESSEE: Patent Counsel
;; STREET: 10666 No. 5652138th Torrey Pines Road, Suite 220,
;; STREET: Mail Drop TPC8
;; CITY: La Jolla
;; STATE: CA
;; COUNTRY: USA
;; ZIP: 92037
;; COMPUTER READABLE FORM:
;; MEDIUM TYPE: Floppy disk
;; COMPUTER: IBM PC compatible
;; OPERATING SYSTEM: PC-DOS/MS-DOS
;; SOFTWARE: Patentin Release #1.0, Version #1.25
;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/08/276,852
;; FILING DATE: 18-JUL-1994
;; CLASSIFICATION: 514
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: US 08/178,302
;; FILING DATE: 30-SEP-1993
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: US 07/954,148
;; FILING DATE: 30-SEP-1992
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Fitting, Thomas
;; REGISTRATION NUMBER: 34,163
;; REFERENCE/DOCKET NUMBER: SCR1452P
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: 619-554-2937
;; TELEFAX: 619-554-6312
;; INFORMATION FOR SEQ ID NO: 170:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 13254 base pairs
;; TYPE: nucleic acid
;; STRANDEDNESS: double
;; TOPOLOGY: circular
;; MOLECULE TYPE: DNA (genomic)
;; US-08-276-852-170

Query Match 60.3%; Score 142.8; DB 1; Length 13254;
Best Local Similarity 81.7%; Pred. No. 3.7e-34;
Matches 165; Conservative 0; Mismatches 37; Indels 0; Gaps 0;
QY 36 GCAGGCATCTAGTGTGTCGAGGACGTGACAGCTAGCCCGGCTCGAGATCTGGGATCTG 95
Db 5179 GCATGATCTCAATTAGTCAGCAACAGGCTCCCGCAGCAGCAAGTATGCAAGCATG 5120
QY 96 CATCTCAATTAGTCAGCAACCATAGTCCCGCCCTAACTCGGCCCATCCGCCCTAACT 155
Db 5119 CATCTCAATTAGTCAGCAACCATAGTCCCGCCCTAACTCGGCCCATCCGCCCTAACT 5060
QY 156 CGGCCAGTTCGGCCATCTCCGCCCATCGCTGACTAATTTTTTTTATTTATGAGAG 215
Db 5059 CGGCCAGTTCGGCCATCTCCGCCCATCGCTGACTAATTTTTTTTATTTATGAGAG 5000
QY 216 GCCGAGGCGGCTCGGCTCTG 237
Db 4999 GCCGAGGCGGCTCGGCTCTG 4978

RESULT 12
US-08-899-575-156
; Sequence 156, Application US/08899575
; Patent No. 5770440
; GENERAL INFORMATION:
; APPLICANT: Burton, Dennis R
; APPLICANT: Barbas, Carlos F
; APPLICANT: Lerner, Richard A
; TITLE OF INVENTION: HUMAN NEUTRALIZING MONOCLONAL ANTIBODIES
; TITLE OF INVENTION: TO HUMAN IMMUNODEFICIENCY VIRUS

TITLE OF INVENTION: TO HUMAN IMMUNODEFICIENCY VIRUS
NUMBER OF SEQUENCES: 170
CORRESPONDENCE ADDRESS:
ADDRESSEE: The Scripps Research Institute, Office of
ADDRESSEE: Patent Counsel
STREET: 10666 No. 5770440th Torrey Pines Road, Suite 220,
STREET: Mail drop TPC8
CITY: La Jolla
STATE: CA
COUNTRY: USA
ZIP: 92037
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/899,575
FILING DATE: 24-JUL-1997
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/276,852
FILING DATE: 18-JUL-1994
APPLICATION NUMBER: US 08/178,302
FILING DATE: 30-SEP-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/954,148
FILING DATE: 30-SEP-1992
ATTORNEY/AGENT INFORMATION:
NAME: Fitting, Thomas
REGISTRATION NUMBER: 34,163
REFERENCE/DOCKET NUMBER: SCRL452P
TELECOMMUNICATION INFORMATION:
TELEPHONE: 619-554-2937
TELEFAX: 619-554-6312
INFORMATION FOR SEQ. ID NO: 170:
SEQUENCE CHARACTERISTICS:
LENGTH: 13254 base pairs
TYPE: nucleic acid
STRANDEDNESS: Double
TOPOLOGY: circular
MOLECULE TYPE: DNA (genomic)
s-08-899-575-170

Query Match	60.3%;	Score 142.8;	DB 1;	Length 13254;	
Best Local Similarity	81.7%;	Pred. No. 3.7e-34;			
Matches 165;	Conservative	0;	Mismatches 37;	Indels 0;	Gaps 0;

36	GCAGGCATCTAGTGTGGTGCAGGACGTGACAGCTAGCCCGGGCTCGAGANVTGCGATCTG	95
2076	GCATGCACTCAATTAGTCAGCAACCAAGGCTCCCGACGAGGCAAGTATGCAAGCAATG	8135
96	CATCTCAATTAGTCAGCAACCATAGTCCGGCCCTAACTCCGCCCATCCGCCCTAACT	155
8136	CATCTCAATTAGTCAGCAACCATAGTCCGGCCCTAACTCCGCCCATCCGCCCTAACT	8195
156	CCGCCCAAGTCCCGCCCATCTCCGCCCCATCGCTGACTAAATTTTTTTTATTTATGCAGAG	215
8196	CGGCCCAAGTCCCGCCCATCTCCGCCCCATCGCTGACTAAATTTTTTTTATTTATGCAGAG	8255
216	GCCGAGGCGCGCTCGGCGCTCTG	237
8256	GCCGAGGCGCGCTCGGCGCTCTG	8277

RESULT 13	
US-08-899-575-170/c	
; Sequence 170, Application US/08899575	
; Patent No. 5770440	
; GENERAL INFORMATION:	
; APPLICANT: Burton, Dennis R	
; APPLICANT: Barbas, Carlos F	
; APPLICANT: Lerner, Richard A	
; APPLICANT: HUMAN NEUTRALIZING MONOCLONAL ANTIBODIES	

RESULT 13
 US-08-899-575-170/c
 ; Sequence 170, Application US/08899575
 ; Patent No. 5770440
 ; GENERAL INFORMATION:
 ; APPLICANT: Dennis R
 ; APPLICANT: Burton,
 ; APPLICANT: Barbas, Carlos F
 ; APPLICANT: Lerner, Richard A
 ; APPLICANT:
 ; HUMAN NEUTRALIZING MONOCLONAL ANTIBODIES